Untangling the tree of life: Which partitioning strategies improve phylogenetic inference?

FELSINGER, STELLA,MARIE

DURHAM UNIVERSITY

MASTER'S THESIS

Untangling the tree of life: Which partitioning strategies improve phylogenetic inference?

Author:
Stella FELSINGER

Supervisors:
Dr Martin SMITH
Dr Richard HOBBS

A thesis submitted in fulfillment of the requirements for the degree of Masters of Science by Research

in the

Palaeoecosystems Research Group
Department of Earth Sciences

June 10, 2020

DURHAM UNIVERSITY

Abstract

Faculty of Science
Department of Earth Sciences

Masters of Science by Research

Untangling the tree of life: Which partitioning strategies improve phylogenetic inference?

by Stella FELSINGER

The rate of evolution is known to vary greatly between morphological characters, which complicates the inference of phylogeny. To accommodate this rate heterogeneity, we can group together characters that are expected to evolve at similar rates. Partitioning by codon position is commonplace in molecular phylogenetics – for morphological data, no such general rule exists.

Rosa, Melo, and Barbeitos (2019) advocate partitioning characters according to their homoplasy index on a maximum parsimony tree. A possible concern with this method is that the choice of tree may influence results in a manner that is not detected by standard model-testing approaches. Such methods rely on a tree to infer a tree – but what if that first tree is unreliable? I tested homoplasy partitioning based on a spectrum of trees ranging from a published tree through to completely random trees, and found that the topology on which homoplasy is calculated does not affect the topology recovered by Bayesian analysis.

I compared homoplasy partitioning to other partitioning strategies, including strategies informed by biological criteria: partitioning by character type (Sereno, 2007) and partitioning by anatomy. Using Steppingstone sampling (Xie et al., 2011) to estimate the fit of each partitioned model, I found that homoplasy is the only reliable approximator of evolutionary rate.

Partitioning has the capacity to significantly increase model fit, but only homoplasy partitioning consistently produced better results than an unpartitioned model. No link between mosaic evolution or character type and evolutionary rates could be confirmed.

Contents

List of Figures

List of Tables

Declaration of Authorship

I, Stella FELSINGER, declare that this thesis titled, "Untangling the tree of life: Which partitioning strategies improve phylogenetic inference?" and the work presented in it are my own.

"Evolution is a light which illuminates all facts, a curve that all lines must follow."

— Pierre Teilhard de Chardin

Acknowledgements

This work was funded by Ørsted A/S and made use of the facilities of the Hamilton HPC Service of Durham University. I benefitted from the expert guidance of my supervisors Martin Smith and Richard Hobbs. Thank you also to my reviewers David Harper and Stuart Jones.

Martin – You have taught me so much over the last three years, and I count myself lucky to have worked with you. Thank you for always making time for me.

Emma – Your honest advice over tea and biscuits was crucial in getting this thesis finished. Thank you so much.

Für meine Eltern

Chapter 1

Introduction

1.1 Introduction

The theory of evolution tells us that any two species of plant, animal, fungus, or bacterium can be traced back to a common ancestor. Understanding how the millions of species alive on Earth today arose is the central challenge of phylogenetics, the statistical framework of biology. Phylogenetic methods illuminate the relationships between species and let us view a species in the context of its relatives. We can compare molecular and morphological information from a group of organisms to build a phylogenetic tree. Well-constrained phylogenies allow us to identify novel pathogens (Lanciotti et al., 1999), discover better medicines by targeting known medicinal plants' closest relatives (Saslis-Lagoudakis et al., 2012; Rønsted et al., 2012), target evolutionarily disparate groups of species for conservation (Polasky et al., 2001), or even infer the behaviour of long-dead animals based on that of their descendants (Field et al., 2018).

As an organism evolves, its appearance and genetic material change. But not all characters change at the same rate. We can account for this rate heterogeneity by partitioning characters into groups that are allowed to evolve under different models of evolution. While molecular data are typically partitioned by gene or by codon position, there is no consensus on which partitioning schemes to use for morphological data. Phylogenetic studies often use only a single partitioning strategy without testing whether it is most suitable for the dataset. The effects of the partitioning strategy on the topology and branch lengths of the reconstructed tree have not been studied on morphological datasets.

In order to find out how best to partition morphological datasets, I tested eight partitioning strategies on five datasets of different animal groups, numbers of characters and taxa, and proportions of fossil taxa. I investigated how variations in the model affected the topology and branch lengths of the inferred tree as well as the fit of the model.

Ultimately, partitioning sheds light on what drives morphological evolution — whether characters that evolve at similar rates are topologically grouped or functionally linked, or whether life stage or character novelty are more influential in deciding optimal partitions.

1.2 The evolution of phylogenetic methods

Phylogenetic trees used to be built by hand on the basis of morphology — the more similar two species look, the more recently they diverged. Modern phylogenetic methods often use molecular data. The steady accumulation of random mutations in a DNA sequence can pinpoint divergence times and increase the resolution of phylogenies at low taxonomic levels. Molecular data allows us to discover cryptic species, who look identical but do not interbreed, or find that a number of very different-looking individuals actually belong to a polymorphic species.

Computational methods for sequence analysis and phylogenetics were ioneered by Margaret Dayhoff in the 1960's and allowed higher throughput analyses to be conducted (Felsenstein, 2001; Dayhoff, 1965), so that trees could incorporate more taxa and more characters. Among the earliest tree selection criteria was Maximum Parsimony, which aims to minimise the net amount of evolution on a tree. It operates on the assumption that the tree requiring the fewest evolutionary changes is the best one. When a character changes more often than necessary, this is attributed to convergent evolution, or homoplasy.

Parsimony remains a popular method for inferring trees, but is today rivalled by Bayesian phylogenetics. Bayesian methods take advantage of modern computing power by testing a large number of parameter combinations and finding the associated trees with the highest likelihood. Bayesian phylogenetics allow the incorporation of prior beliefs about evolution and hypotheses on evolutionary patterns. The most popular program for Bayesian interference, MrBayes (Ronquist et al., 2012), models morphological character evolution with the Mk model (Lewis, 2001). This model effectively treats morphological characters like bases in a DNA sequence. This is an imperfect approximation of their actual evolutionary behaviour, mainly due to the strong heterogeneity in evolutionary rates among morphological characters. A popular approach to improving the Mk model's fit is to partition the characters into subsets thought to share similar evolutionary patterns.

1.3 Successful partitioning

1.3.1 Formulating morphological characters

Molecular characters all tend to follow the same statement scheme. For example, DNA or RNA sequences consist of individual nucleotides. Each position in a sequence can be considered a character with four possible states corresponding to the four bases A, T, C, and G. In contrast, morphological characters are much more heterogeneous. It is challenging to condense an organism's physiology into a collection of statements on the presence, absence, or state of isolated features — to do so in a way that can be generalised to a larger group of taxa is very difficult indeed.

Comprehensive frameworks exist to guide taxonomists in the process of character selection and coding, and there is a need for more authors to follow them in the

interest of establishing best practice and simplifying collaboration. Brazeau (2011) discusses the problematic consequences of several shortcomings of character coding methods and gives advice on recoding problematic characters. Sereno (2007) provides a rigorous framework for defining characters to better illustrate character dependencies and hierarchies, using a division of characters into two types — neomorphic characters which have only presence and absence states, and transformational characters which describe quantitative or qualitative states of a present feature. Brazeau, Guillerme, and Smith (2019) make recommendations on treating inapplicable characters, which unavoidably arise when characters are formulated rigorously according to Sereno (2007), and Vogt (2018) discusses the coding of ontologically dependent characters.

1.3.2 What makes a good partitioning strategy?

A number of strategies for partitioning morphological data have been proposed and employed, including partitioning characters by anatomy (Clarke and Middleton, 2008), by developmental stage (Tarasov and Génier, 2015), by character type (Sun et al., 2018b; Moysiuk and Caron, 2019), and by homoplasy on a maximum parsimony tree (Rosa, Melo, and Barbeitos, 2019).

Most of these partitioning strategies rely heavily on the author's biological judgement. The author must decide, usually by carefully surveying the character list or based on a preformed hypothesis about the evolution of a group, which partitions to establish, and then assign each character to a partition. Which partition a character suits best is often open to interpretation. Characters may have been defined with a specific partitioning scheme in mind, and fitting them into a different scheme can then be difficult. Aside from ensuring that partitions reflect biologically sensible groupings, the author must also take into account the size of each partition and the overall number of partitions, as these will influence the success of parameter estimation and the suitability of the model (Holder and Lewis, 2003).

A good partitioning strategy must be reproducible — two independent researchers partitioning the same dataset should end up with the same partitions. Ideally, the scheme should be applicable by someone without deep familiarity with the taxonomic group. No specialist knowledge of their lifecycle or physiology should be required, and where necessary an explanation of the character should be provided by the author of the character list.

One solution is to compare results from as wide a range of partitioning schemes as possible. The program PartitionFinder2 accomplishes this heuristic search for an optimal partitioning strategy (Lanfear et al., 2012; Lanfear et al., 2016). Since there are too many possible partitioning schemes to directly compare them all, PF2's search is based on mixing and matching user-defined blocks of characters into subsets. The partitioning scheme comprising the most suitable subsets is then selected by an information-theoretic approach using each model's maximum likelihood.

1.3.3 The effects of partitioning

Partitioning allows us to model differences in evolutionary behaviour across characters more effectively. The estimated parameters include branch lengths, the rates of changing character state, and the topology of the tree. If there are large differences in evolutionary rate between characters, a single parameter estimate for the rate could be insufficient. Estimating a rate distribution for each partition can improve inference compared to an unpartitioned analysis. The partitioned model should better emulate the real evolutionary processes and receive a higher measure of model fit (Marginal Likelihood, see section 3.1.3.1: the likelihood of the data given the model averaged over all parameters), unless the added effort of estimating more parameters outweighs the benefit of improved inference.

Aside from potentially improving model fit, the further effects of partitioning have not been explored sufficiently. Most efforts have concerned molecular datasets. Brown and Lemmon (2007) emphasized the importance of partitioning large datasets and demonstrated that models with too few partitions lend support to the wrong trees. Strugnell et al. (2005) showed that partitioning strategy choice affects node support. Analyses by Duchêne et al. (2011) showed that six mitochondrial genes in various combinations inferred trees which differed in node support, branch lengths, and topology. Poux et al. (2008) analysed a molecular dataset of Malagasy tenrecs and found that models with more partitions estimated older node ages. Standard deviations of node ages were highest for the minimum and maximum number of partitions, indicating that an intermediate number of partitions (in their case 5 and 9) allowed node ages to be estimated with the highest confidence. In contrast, some studies found no significant effects of partitioning strategy choice on the inferred tree, e.g. Cameron et al. (2012). In a comparative study of 34 molecular datasets, Kainer and Lanfear (2015) concluded that topology is highly affected by partitioning strategy, as are node support and branch lengths to a lesser degree. They suggest also that implementing too few partitions has a greater capacity to lead to bad results than splitting the data over too many partitions, supporting the results of Brown and Lemmon (2007).

Judging by the often strong effects of partitioning strategy on tree characteristics in molecular datasets, it is likely that the same problems occur in morphological datasets. Mickevich (1978) found that estimates of topology were similar when analysing data partitions separately using Wagner parsimony, but these analyses included molecular as well as morphological data. Benson (2012) on the other hand found that different anatomical partitions reflected different topologies in a morphological analysis of basal synapsids.

To make accurate reconstructions of a clade's history it is necessary to be aware of the possible implications of choosing a suboptimal partitioning strategy. There is currently a lack of knowledge about these implications as well as what might constitute a suitable partitioning scheme. Thus it remains necessary to further study

the effects of partitioning scheme choice on aspects of the inferred tree, such as branch lengths and topology.

1.4 My approach

I used Bayesian inference to estimate posterior trees along with associated parameter values and measures of model fit. Five datasets from the literature were partitioned according to eight different partitioning schemes. I also employed four branch length priors including a relaxed morphological clock model.

By analysing five empirical datasets under a wide range of partitioning schemes, I attempt to answer the following questions:

- How does the partitioning scheme influence tree topology and branch lengths?
- Which partitioning strategies and branch length priors produce the best-fitting models?
- What can partitioning tell us about morphological evolution?

1.5 Roadmap

In the Background chapter I outline current methods in computational phylogenetics and give an introduction to Bayesian inference methods.

In the Methods, I describe the datasets, branch length priors, and partitioning strategies I compared. I give instructions for partitioning morphological data, explain how to set up and conduct phylogenetic analyses in MrBayes, reference relevant scripts (provided online, see section 3.7.3) and software recommendations, and detail how results were analysed.

In Results I, I present the effects of varying the partitioning tree (see section 3.1.2.2) on model fit, branch lengths and tree topology under homoplasy partitioning. I also investigate the relationship between homoplasy and Information Content of a character. Results II compares the model fit of analyses under eight partitioning strategies and four branch length priors.

Finally I discuss which partitioning strategies show promise, and make recommendations on the use of partitioning to other workers in phylogenetics. I also suggest future avenues for research into partitioning morphological data.

Chapter 2

Background

In this chapter, I first outline the philosophical underpinnings of phylogenetics from ancient history of representing biological interrelationships in the form of trees of descent. I discuss how the understanding of evolution and natural selection evolved and how the earliest phylogenetic studies were conducted. I then introduce the beginnings of computational phylogenetics in the wake of the Modern Synthesis. I follow on to talk about the advantages and disadvantages of molecular and morphological data. The last section introduces the principles of Bayesian phylogenetics and discusses the origins of partitioning data.

2.1 A (very) short history of phylogenetics

2.1.1 The Tree of Life

Phylogenetic trees are the overarching organising principle in natural history. Darwin's Origin of Species contains what is often considered the first phylogenetic tree, but the connection of life and trees has much deeper roots — the world tree Yggdrasil in Norse mythology for example, or the Tree of Knowledge in the Christian creation myth, associating knowledge, order and logic with a tree. Creation myths almost universally describe the beginning of our world as order emerging from chaos, the divine entity being likened to the ordering principle. The idea that our world is inherently one of order means that we should be able to find and describe that order, and place everything in existence into this logically sound framework. This is what early scholars attempted through the use of trees (Wilkins, 2009).

Ancient Greek scholars were particularly interested in the constancy of words over matter. Their early attempts at classification included the observation that matter changes over time and yet remains the same — humans age, but stay human. This branch of logic formulated the concepts of genus and species (Wilkins, 2009). These terms were originally used simply to group and distinguish objects and ideas and did not refer to organisms in any evolutionary sense, as the Ancient Greek view of nature did not foresee change in the essence of a thing; A lioness would change throughout its lifecycle, but would never alter the substance of her being sufficiently to become anything other than a lioness (Gontier, 2011).

Ancient Greek philosophy furthermore laid the foundation of what we nowadays would refer to as characters. Aristotle defined species of animate as well as inanimate objects by what he called 'differentia': qualities which one species possessed and another lacked (Deverreux and Pellegrin, 1990). The possession of a 'soul', that differentium which distinguishes living from lifeless matter, was of particular interest to Aristotle. He viewed plants as intermediate on the scale from inanimate to animate, reasoning that they were less alive than animals but more alive than, for example, rocks.

A very early tree diagram is the Arbor Porphyriana, or Tree of Porphyry, named after the influential Neo-Platonist. This diagram placing the human species into the context of all matter, animate and inanimate, dates back to the sixth century (Verboon, 2014). The Tree of Porphyry was reproduced hundreds of times throughout the Middle Ages and distributed across all of Europe, influencing logical thought for millenia to come (Ragan, 2009; Gontier, 2011). Tree diagrams were used to visualise lines of descent in family trees, or degrees of relatedness among languages in linguistic pedigrees (Schleicher, 1853). It follows that biologists attempted to order the natural world using this well-established framework. The structure which had served to organise everything from clerical hierarchies to racehorse pedigrees was imposed on species relationships too. When the growth direction of the tree came to represent time, the phylogenetic tree as we know it today was in existence (Wilkins, 2009).

2.1.2 The beginnings of phylogenetics in actual fact

Advances in Geology and Palaeontology in the 18th century led to two discoveries:

1. The Earth was much older than previously believed.

2. Animals had existed in the past which showed significant morphological differences to their living counterparts, or did not resemble any known creature.

Coupled with an immense time scale to operate on, these discoveries established extinction as a fact of life and paved the way to understanding how origination and extinction could operate as driving forces in an evolutionary process (Gontier, 2011; Ragan, 2009).

Early phylogeneticists placed taxa onto trees using the only data available to them — physical features, or morphology. Similarities were used to group taxa together and dissimilarities to distinguish groups, just as Aristotle had divided the world into inanimate and animate substances by their differentia. This way of ordering taxa was heavily based on structures with the same apparent build or function indicating affinity among the organisms which possess them.

Early phylogenetic studies like Edwards and Cavalli-Sforza (1963) used the principle of parsimony, the idea that evolutionary change would take the shortest route or the least effort to happen. Trees with the fewest changes were favoured

over those requiring more changes in a character. These early studies were limited by the processing power of the available computers; Exhaustive pairwise comparison of a group grew more complicated and time-intensive the larger the number of taxa involved. As time went on, more computationally intensive methods became available, notably Bayesian methods which require nontrivial integrations over multidimensional parameter space.

2.1.3 The Molecular Revolution

In 1869, Friedrich Miescher discovered what would come to be known as DNA (Dahm, 2008). This monumental discovery began a chain of studies into the function and structure of this new molecule crowned by Rosalind Franklin and Maurice Wilkin's work on X-ray crystallography, which led to the discovery of the double-helix structure of DNA molecules in 1953 (Klug, 1968). Dayhoff's work on computational protein sequencing and Sanger's work on gene sequencing methods (Dayhoff, 1965; Sanger and Coulson, 1975; Sanger, Nicklen, and Coulson, 1977) were soon expanded upon to yield high-throughput massively parallel gene sequencing technologies (Moorthie, Mattocks, and Wright, 2011).

Now, it was possible to reduce an organism, as complex as it might be, to a string of letters. The similarity between the strings of two organisms was a direct, inambiguous measure of their affinity. Once the genetic code of any organism could be read with relative ease, this became the major avenue of phylogenetics, overtaking morphological methods. With growing evidence that fossils contribute vital data to phylogenies, however, morphological data is enjoying a renaissance, and studies increasingly attempt to integrate insights from several data types (Arcila et al., 2015).

2.2 Molecular and morphological data

Morphological data has a reputation for being difficult to work with. Indeed there are some challenges peculiar to morphological characters, but there are difficulties with molecular data too: Alignment issues (Löytynoja and Goldman, 2008), variability between individuals (Göker and Grimm, 2008), and single-site polymorphisms (Caccone et al., 1996) are common and can confound molecular phylogenetic analyses. Each type of data has its challenges and shortcomings, as well as great potential (Wiens, 2000). There is phylogenetically informative and uninformative variability in molecular sequences, just as in morphological data. As we will see, both types of data are valuable in the reconstruction of phylogenetic relationships.

2.2.1 Advantages of molecular data

Molecular sequencing offers a massive number of observable characters. The number of characters can go into the tens of thousands, e.g. Kawahara et al. (2017). Shorter 'genetic barcodes' are also often used, like the chloroplast gene *rcbL* (Janzen, 2009) or portions of the ribosomal subunits (Wu, Xiong, and Yu, 2015). A large number of characters can help to successfully estimate parameters and improve accuracy of the reconstructed tree, though not as effectively as the addition of taxa (Graybeal, 1998).

In morphological data, there is a natural limit on the number of observable independent characters. A massive mammal dataset of 4541 morphological characters was compiled and analysed under parsimony by O'Leary et al. (2013). The resulting tree was subsequently found to be highly influenced by homoplasy, resolving functional groups rather than phylogenetic lineages and requiring unnaturally fast rates of nucleotide substitution (Springer et al., 2013). This indicates that as the number of morphological characters increases, the risk of sampling dependent characters and characters with high homoplasy rises as well.

In the absence of gene linkage and polynucleotide substitutions, individual sites of a molecular sequence behave as independent characters, which is a requirement of most phylogenetic algorithms. Non-independence is usually detected when producing a sequence alignment (for example in the case of polynucleotide deletions or insertions), and can be accounted for by reducing the weights assigned to dependent characters (Chippindale and Wiens, 1994).

Modern methods make collecting molecular data faster and easier than collecting morphological data. The phylogeneticist must only choose a gene to sequence. Protocols and sophisticated machines then simplify and automate the processes of DNA extraction and amplification, and publicly available algorithms like BLAST take care of aligning sequences.

Evolutionary patterns of molecular data are quite well known. The rate of change for DNA is known to vary between genes and between regions of a gene (e.g. protein-coding exons and non-coding introns). Changes to DNA accumulate at a fairly constant rate and are not strongly influenced by selection, as long as they are functionally neutral (Kimura, 1983). The number of base differences in a given gene from two species can thus be considered proportional to the time since their divergence (the "Molecular Clock", see Morgan (1998) and Dos Reis, Donoghue, and Yang (2016)).

2.2.2 Advantages of morphological data

Molecular data is only available for extant taxa and the most recent of fossils. By constrast, even heavily decayed fossils offer up some morphological data. Morphology allows us to place fossils from deep time, museum specimens with denatured DNA, and taxa from inaccessible localities, on trees.

The inclusion of fossils can help to break up long branches between widely diverged clades, avoiding long branch attraction (Huelsenbeck, 1991). Long branches can lead to wrong clades receiving strong support (Huelsenbeck, 1995; Felsenstein, 1978a; Huelsenbeck, 1997). Fossils can also provide a root for the tree and help date interior nodes of the tree via earliest appearances, as with the Stratigraphic Consistency Index (Huelsenbeck, 1994) or calibrating divergence times (Dos Reis, Donoghue, and Yang, 2016). They can also help to resolve relationships between extant taxa (Gauthier, Kluge, and Rowe, 1988).

Even incomplete fossils are worth including, as missing data does not necessarily lead to poorly resolved trees (Kearney and Clark, 2003). Incomplete fossil taxa can also contribute important information by showing which states were ancestral (Huelsenbeck, 1991).

Morphological data collection is much cheaper than sequencing DNA for potentially hundreds of taxa. Morphometric data collection can even be crowdsourced without loss of efficacy (Chang and Alfaro, 2016). The coding of morphological character data from images is usually performed by scientists, but crowdsourced workers fulfilled the task similarly well. While not every type of character can be coded from an image, this method of gathering data nevertheless has great potential in allowing the size of morphological datasets to keep pace with molecular datasets (Parins-Fukuchi, 2018).

If the cost of adding a new taxon is small, the dataset can include data on more taxa. Graybeal (1998) showed that it is more beneficial to add taxa rather than characters in order to resolve problematic phylogenies. She analysed a total of 40,000 nucleotides for groups of taxa ranging in number from 4 to 30, splitting the nucleotides over the number of taxa evenly. Even as each taxon was represented by fewer characters, the proportion of correctly reconstructed branches increased as more taxa were incorporated into the tree, demonstrating the importance of broad taxon coverage.

Morphological evolution is always the phenotypical expression of molecular evolution, but there is rarely a one-to-one link between genotype and phenotype. Several mutations can cause the same phenotype because most morphological characters are linked to several genes (Manceau et al., 2010). On the flipside, complex morphologies can be controlled by just a few genes (Schaap et al., 2006). While this creates the possibility of mistaking several independent mutations for one shared mutation, a morphological phylogeny cannot be misled by a strongly supported gene phylogeny which does not reflect the true species phylogeny.

Morphological evidence can be used to verify a molecular phylogeny. If it is not supported by morphology, it can be enlightening to find out why the two sources of evidence disagree (Lee and Palci, 2015). Discord between palaeontological and molecular evidence ("Rocks and Clocks") can be caused by non-independence of morphological characters, inflating branch lengths and leading to significant ghost

lineages through the estimation of divergence times that far predate first appearances in the fossil record ('Deep Root Attraction', Ronquist, Lartillot, and Phillips (2016)). Deep Root Attraction can also be caused by vague or misspecified priors and unsuitable models. Reanalysis of the problematic supermatrix of O'Leary et al. (2013) (recall section 2.2.1) in combination with molecular data by Ronquist, Lartillot, and Phillips (2016) showed that inference was improved under more informative priors and more appropriate evolutionary models for both data types. Morphological characters also have irreplaceable value to phylogenetics because only they offer insight into deepest time where DNA sequences are saturated with changes — new mutations simply override previous mutations, so that the amount of molecular evolution that has taken place is underestimated.

Finally, it is worth remembering that morphology remains our principal way of recognizing species. Phylogeneticists sampling molecular diversity usually do so on the basis of species identifications which were diagnosed with morphology.

2.3 Background on Bayesian Phylogenetics

Bayesian and other likelihood-based methods model evolution explicitly by estimating the timing and frequency of changes of state in each character.

A Bayesian approach requires the setting of priors, which describe our beliefs about the evolutionary history of a clade and allow us to incorporate this information into the model. It is necessary to set a prior on any estimated parameter including branch lengths, topology, and transition rates between character states. This allows us to quantify the uncertainty both in our prior beliefs as well as in the resulting parameter estimates (Holder and Lewis, 2003). The combined prior is multiplied with a likelihood function approximated from the data to result in the posterior probability distribution, which assigns each possible tree the likelihood that it reflects the true phylogeny. This is computationally intensive, but tractable through techniques which approximate the posterior probability distribution through gradual exploration (e.g. hill-climbing algorithms) rather than analytically through computation.

Contrary to other likelihood-based methods, Bayesian models are usually compared not by the likelihood of the best tree (Maximum Likelihood), but by the average fit of the model after all marginal parameters have been integrated out — the Marginal Likelihood (see 2.3.2.1). The surface under the posterior probability surface can be considered equivalent to the support of the data for the evolutionary model, also referred to as model fit.

By varying the priors and parameter settings of a Bayesian model, we can emulate different evolutionary hypotheses and directly compare their model fit. We can tell which of hypothesis is most likely to reflect biological truth, but cannot predict whether another hypothesis might be more accurate. For example, we could test the hypothesis that euarthropods and onychophorans (Arthropoda) are more

closely related than euarthropods and tardigrades (Tactopoda). After specifying a topology constraint enforcing the grouping of either Arthropoda or Tactopoda and including it as the topology prior, we can estimate the marginal likelihood of both models and get an indication of which hypothesis is more accurate. But this model comparison cannot tell us if both of our initial hypotheses are wrong, and onychophorans and tardigrades are actually more closely related to each other than either is to arthropods.

Controversy surrounds the question whether Maximum Parsimony or Bayesian methods best model morphological evolution. A detailed comparison is beyond the scope of this thesis, so it shall suffice to say that parsimony can perform well as long as branches are short and do not vary too much in length (Holder and Lewis, 2003). Parsimony and Bayesian methods are regularly used in the same study to verify results (e.g. Potter et al. (2007), Tarasov and Génier (2015), and Sun et al. (2018b)). Implied weights parsimony specifically has been identified as performing as well as Bayesian analysis on simulated datasets (Smith, 2019a). However, Bayesian models can be used to test and compare different hypotheses of evolution, which is not possible using parsimony approaches.

2.3.1 The Mk model (Lewis, 2001)

The popular phylogenetics program MrBayes (Ronquist et al., 2012) uses the Mk model (where "M" stands for a Markov chain modelling a character with k observed states) as specified by Lewis (2001) to infer phylogenies from discrete morphological characters. The Mk model is closely based on the Jukes and Cantor model of protein evolution which uses Bayesian Markov chain Monte Carlo (Jukes and Cantor, 1969). Markov chain Monte Carlo is a stochastic simulation method that allows us to numerically approximate probability distributions which are too large to be computationally evaluated (Geyer, 1992). It employs Markov chains which traverse the probability distribution, moving from state to state and taking regular samples which together form an estimate of the target distribution.

In a phylogenetic context, the Mk model is applied to estimate the posterior distribution of phylogenetic trees with their associated probabilities of being the true tree. The probability of whether or not the chain accepts a proposed state depends on the likelihood of the current tree: a chain is more likely to change state if the current tree is “bad”, i.e. unlikely to reflect the true phylogeny.

Bayesian phylogenetics offers many intractable problems due to the sheer number of phylogenetic trees that must be evaluated (Felsenstein, 1978b). Even just the number of different tree topologies $T(n)$[1] grows factorially with the number of taxa (Felsenstein, 2001):

$$T(n) = \frac{(2n-3)!}{2^{n-2}(n-2)!} \tag{2.1}$$

[1] $T(n)$ here represents the number of possible rooted bifurcating labelled trees with n terminals.

In addition, branch lengths can vary considerably between different phylogenetic trees. Branch lengths represent the passage of time between divergence events and are tightly linked to the evolutionary rates, as they are proportional to the number of changes of state along the branch (Ronquist, Huelsenbeck, and Teslenko, 2011). Branch length is defined as $(k-1)\alpha t$, where k = the number of states of a character, α = the instantaneous rate of a state transition (as defined in the rate transition matrix), and t = the length of time represented by that branch. The instantaneous rate matrix Q provides a mapping between states and holds the rates of transitioning from one to the other in any given instant (Lewis, 2001):

$$Q = \alpha \begin{bmatrix} 1-k & 1 & ... & 1 \\ 1 & 1-k & ... & 1 \\ ... & ... & ... & ... \\ 1 & 1 & ... & 1-k \end{bmatrix} \tag{2.2}$$

2.3.1.1 Estimating the probability of a single tree (adapted from Huelsenbeck and Ronquist (2001))

Calculating the posterior probability distribution for the whole of tree space is computationally infeasible. The posterior probability of a single tree τ_i given the data X can be expressed using Bayes' theorem:

$$f(\tau_i|D) = \frac{f(D|\tau_i)f(\tau_i)}{\sum_{j=1}^{T(n)} f(D|\tau_j)f(\tau_j)} \tag{2.3}$$

where

$$f(D|\tau_i) = \int_v \int_\theta f(D|\tau_i, v, \theta) f(v, \theta) dv d\theta. \tag{2.4}$$

and

$\tau_i =$ the ith phylogenetic tree topology

$D =$ the data

$T(n) =$ the number of possible trees

$\theta =$ the substitution parameters

$v =$ the branch lengths

$f(\tau_i|D)$ depends on the sum of the probability of the data given the tree ($f(D|\tau_i)$) for all possible trees. This involves a computationally intractable double integral over branch lengths v and substitution parameters θ. D represents an alignment of molecular or morphological character data. $f(\tau_i)$ designates the tree topology prior, while the prior for branch lengths and substitution parameters is $f(v, \theta)$.

2.3.1.2 Exploring tree space

A Markovian sampler moves through tree space, collecting the current tree and its location in parameter space at a given sampling frequency.

A basic MCMC algorithm consists of three repeated steps (Huelsenbeck and Ronquist, 2001):

1. A new state for the chain is proposed. This can be a change in tree topology, in a parameter, or in a branch length. These changes are accomplished via a range of moves in MrBayes, which each have varying probabilities of being proposed.

2. The probability with which this new state is accepted over the current state is calculated. This acceptance probability is the product of 3 components:

the likelihood ratio	the ratio between the likelihoods of the current and proposed state, or 1, whichever is smaller
the prior ratio	the ratio between the priors of the current and proposed state
the proposal ratio	the ratio between the probability of proposing the current and proposed state

3. A random value between 0 and 1 is drawn. If this value is lower than the acceptance probability of the new state, the chain accepts the new state.

Every iteration of these three steps constitutes a generation. If run for enough generations, the samples drawn from the chain represent a valid sample of the target distribution of interest. When the sampler has explored tree space sufficiently, the number of times a given tree is sampled is proportional to its posterior probability. Using all sampled trees, we can calculate an estimate of the high-dimensional integral and the posterior probability distribution from the beginning of this section.

Depending on the size of the posterior distribution, it can be difficult for just one Markov chain to reach all regions. Running multiple chains in parallel can improve the estimate of the target distribution. But in rugged tree space, there are often local peaks surrounded by areas of low likelihood which chains are unlikely to traverse (Brown and Thomson, 2018). Metropolis-coupled MCMC or MC^3 is a modification of the standard Markov chain Monte Carlo algorithm which attenuates this behaviour (Altekar et al., 2004). In addition to an unmodified "cold" chain, MC^3 runs a number of "heated" chains which are able to traverse tree space more easily. Samples are drawn only from the cold chain, while each heated chain explores a different transformation of the posterior distribution with flattened peaks and shallower valleys. The heated chains regularly swap places with adjacent chains of different temperatures. This lets the conservative cold chain jump between regions of the posterior to produce a more comprehensive sample (Huelsenbeck and Ronquist, 2001).

The generalised Mk model makes no assumptions regarding plesiomorphy or apomorphy of any character state. Changes occur during "instants", defined as infinitesimal periods of time dt, and occur with symmetrical probability, meaning that changing from state i to j is as probable as the reverse (Lewis, 2001). I modified this in the case of neomorphic-transformational partitioning, where I specified a higher likelihood of gaining the "present" state than losing it.

2.3.2 Steppingstone sampling (Xie et al., 2011)

The following section is heavily based on the methods section of Xie et al. (2011) and on Fourment et al. (2019).

The aim of Steppingstone sampling is the approximation of a model's marginal likelihood, the central unit of Bayesian model comparison, via importance sampling. The approach involves defining a path of distributions connecting the prior and posterior. Each intermediate distribution is approximated using MC^3 and stepwise likelihood ratios are estimated, allowing the crossing of the gulf between prior and posterior as though via stepping stones. The product of the intermediate likelihood ratios is equal to the marginal likelihood, as long as the priors are proper.

Steppingstone sampling approximates marginal likelihoods more accurately than the harmonic mean method, which tends towards overestimation (Xie et al., 2011). The harmonic mean of the marginal likelihood can be easily calculated after an MC^3 analysis and is included in the standard output in MrBayes, but in my experience these estimates show too much variability to compare multiple models, some of which might not show large differences in performance.

2.3.2.1 The Marginal Likelihood

The marginal likelihood is the probability of the data given the model averaged over the parameter space and taking into account the prior (adapted from (Oaks et al., 2019)):

$$p(D|M) = \int p(D|\theta, M)p(\theta|M)d\theta \tag{2.5}$$

where

$D =$ the data

$\theta =$ the parameter vector $(\theta_1, \theta_2, ..., \theta_N)$

$M =$ the model

$p(\theta|M) =$ the prior on θ

$p(D|\theta, M) =$ the probability of the data given the vector of parameters θ

2.3.2.2 Importance sampling

Importance sampling can accelerate Monte Carlo processes by increasing sampling of areas of parameter space that are thought to have the largest impact on the marginal likelihood (Rubino and Tuffin, 2009). An importance distribution is used which is similar to the target posterior distribution. To avoid biasing the result, each sampled value is normalised by the probability of sampling that value from the importance distribution, its importance weight. A definition of importance sampling can be found in Oaks et al. (2019).

In Steppingstone sampling, the estimated posterior from the previous step serves as the importance distribution for the current step.

2.3.2.3 Steppingstone sampling

Let us define a power posterior density function $q_\beta(\theta)$ that is normalised by the power marginal likelihood c_β to yield p_β, the normalised power posterior density:

$$\begin{aligned} q_\beta &= f(D|\theta, M)^\beta f(\theta|M) \\ p_\beta &= \frac{q_\beta}{c_\beta} \end{aligned} \tag{2.6}$$

q_β is the product of the likelihood function $f(D|\theta, M)$ and the prior $f(\theta|M)$.

When $\beta = 1.0$, the power posterior q_β is equivalent to the posterior distribution. When $\beta = 0.0$, q_β is equivalent to the prior distribution.

If the prior is proper, $c_{0.0} = 1.0$ and the marginal likelihood is equivalent to the ratio $r_{ss} = \frac{c_{1.0}}{c_{0.0}} = c_{1.0}$. This central ratio can be approximated as a product of K stepwise ratios:

$$\begin{aligned} r_{SS} &= \frac{c_{1.0}}{c_{0.0}} \\ &= \prod_{k=1}^{K} \frac{c_{\beta_k}}{c_{\beta_{k-1}}} \\ &= \prod_{k=1}^{K} r_{SS,k} \end{aligned} \tag{2.7}$$

where $r_{SS,k} = \frac{c_{\beta_k}}{c_{\beta_{k-1}}}$ for step $k \in [1, K]$.

Each ratio $\frac{c_{\beta_k}}{c_{\beta_{k-1}}}$ is estimated consecutively by importance sampling with $p_{\beta_{k-1}}$, the power posterior estimated in the previous step, as the importance distribution. This works because $p_{\beta_{k-1}}$ is only slightly closer to the prior, i.e. slightly more dispersed, than the power posterior distribution of the current step, and thus serves as a good approximation.

The k^{th} ratio expressed using importance sampling is

$$\begin{aligned}\hat{r}_{SS,k} &= \frac{1}{n}\sum_{i=1}^{n}\frac{p(D|\theta_{k-1,i})^{\beta_k}}{p(D|\theta_{k-1,i})^{\beta}_{k-1}} \\ &= \frac{1}{n}\sum_{i=1}^{n} p(D|\theta_{k-1,i})^{b_k - b_{k-1}}\end{aligned} \tag{2.8}$$

where $\theta_{k-1,i}$ is a MCMC sample vector from $p_{\beta_{k-1}}$, and $f(D|\theta_{k-1,i})$ is the likelihood of that sampled parameter vector.

The product of these K ratios is a faithful estimate of the true marginal likelihood of the target distribution:

$$\hat{p}_{SS} = \prod_{k=1}^{K}\hat{r}_{SS,k}. \tag{2.9}$$

2.3.3 Partitioning and weighting

Partitioning originated as a way of dealing with diverse datasets. As early as the 1970s, Mickevich (1978) used partitioning to compare trees from different subsets of data. He separated types of morphology — such as male and female bee characters in the Hoplites dataset of Sokal and Michener (1967)), morphological and allelic data (kangaroo rats (Johnson and Selander, 1971; Schnell, Best, and Kennedy, 1978)), or types of molecular data (Amino acid sequence of primate Alpha and Beta hemoglobin (Dayhoff, 1969) — into character subsets. Mickevich computed a tree for each subset and built a consensus tree from the shared topological features in a method he called 'Taxonomic Congruence'. It became apparent that different sets of characters suggested different phylogenies (see also Rodrigo et al. (1993)). Taxonomic Congruence was also used by Bielecki et al. (2014), who compared trees of leech phylogeny reconstructed from a set of 22 characters relating to female reproductive and oocyte morphology, as well as from a more inclusive dataset including these and 27 general morphological characters.

The approach of Mickevich (1978) was countered by Kluge (1989), who argued for combined analysis of all available data ('Total Evidence'). Bull et al. (1993) and De Queiroz (1993) suggested an intermediate method named 'Prior Agreement' whereby each partition is analysed separately and trees are examined for agreement. If partitions are found to disagree, the cause of the heterogeneity is identified. Once all partitions are in agreement, the data is reanalysed in combination. Derome et al. (2002) state that they did not follow the Prior Agreement approach of Bull et al. (1993), but they did perform separate phylogenetic analysis on each subset (Cytochrome b gene, d loop region, and morphological characters) before estimating a tree from the combined data. They note that despite the much greater number of molecular to morphological characters (744:298:26), the combined tree topology was most influenced by the morphological characters.

Separate analysis is sometimes advocated when data types underlie different evolutionary processes, such as morphological and molecular data. When the method of reconstruction assumes a specific evolutionary model, lumping different types together might not lead to the best results (Bull et al., 1993). Chippindale and Wiens (1994) argue that these differences in evolutionary process can be accommodated and accounted for using weighting. Weighting characters differently can even account for character dependence. Most weighting schemes weight characters by homoplasy, a measure of how often a character changes on a tree. Assuming that evolution would have taken the shortest path, characters with few changes are afforded more weight and the weight of apparently homoplastic characters is reduced. Weighting functions are typically concave, so that extra steps on already highly homoplastic characters decrease their weight by only a little, while extra steps on characters without homoplasy drastically decrease their weight. Establishing the homoplasy of a character requires a tree, so these methods often repeatedly analyse the same data (e.g. Successive Weighting, Farris (1969)). Goloboff (2013) proposed weighting entire partitions collectively in parsimony analyses, for example weighting third codon positions by their average homoplasy. Because both methods allow segregation of characters according to some criterion, weighting can be considered the parsimony equivalent to partitioning in Bayesian analyses.

Nylander et al. (2004) used partitioning for the first time in a Bayesian approach, referencing the work of Yang (1996) on accommodating rate variation under Maximum Likelihood. Partitioning has become an integral part of many Bayesian phylogenetic studies (e.g. Brandley, Schmitz, and Reeder (2005), McGuire et al. (2007), and Lee and Worthy (2012); see also section 2.3.3.2). This method can be considered a type of total evidence approach. Partitioning is not a step prior to analysis, but rather is a part of the analytical process and the model. It provides information to the model about the expected parameters governing the evolutionary process of each set of characters. Branch lengths can estimated separately for each partition, but typically only a single topology is inferred based on the entire dataset (Rosa, Melo, and Barbeitos, 2019). Ronquist, Lartillot, and Phillips (2016) advocate a Total Evidence approach to estimating divergence times and explain why molecular, morphological, and palaeontological evidence may disagree.

2.3.3.1 Examples of molecular partitioning

The genetic code is redundant, meaning that several three-base codons code for the same amino acid. Often the difference between these synonymous codons lies in the third base. A mutation in this position is functionally neutral as it does not change the amino acid the codon is translated into. Mutations in this "wobbly" third base (Crick, 1966) are thus less likely to have a phenotypic expression which natural selection could act on. This suggests that third codon positions are less strictly conserved than the others.

In a study by Hillis, Huelsenbeck, and Swofford (1994), parsimony performed better than some model-based alternatives when characters with high frequencies of change are downweighted. Buckley (2002) emphasised the importance of partitioning in allowing for rate heterogeneity among characters. He implicated overly simplistic models in generating overconfident estimates of posterior probabilities and stressed the importance of correctly specifying evolutionary models for maximum likelihood and Bayesian analysis. Brandley, Schmitz, and Reeder (2005) analysed a 2200 base sequence of mitochondrial DNA under a range of partitioning strategies using Bayesian inference. They found that the success of a partitioning strategy relied not on the number of partitions, but more on the partitioning criterion. Bofkin and Goldman (2007) tested a model with three partitions based on codon position using maximum likelihood and found significant differences in the preferred evolutionary models for each position. Evolutionary rates in the third codon position were significantly higher and differently distributed than rates in the second position. But faster rates of change in third positions do not mean they are without value for molecular phylogenetics. Källersjö, Albert, and Farris (1999) demonstrated that the exclusion of third positions decreased the support a number of widely accepted monophyletic groups received. Rather than excluding these characters, it is sufficient to account for their peculiarity by assigning them different weights than the other characters. In an analysis by Goloboff et al. (2008), downweighting rather than exclusion of third positions increased support for several groups and produced phylogenies that were stable under addition of further characters and taxa. Strugnell et al. (2005) also found that partitioning by codon position significantly improved results of a Bayesian phylogenetic analysis compared to not partitioning the data.

2.3.3.2 Examples of morphological partitioning

Anatomy — Gaubert et al. (2005) used partitioning by anatomy to study patterns of adaptive convergence in feliformian mammals. Clarke and Middleton (2008) found that anatomical partitioning improved inference of phylogenetic relationships between crown-group birds in the study of the early evolution of flight. The same partitioning method was also employed by Tarasov and Génier (2015), who called it anatomy ontology.

Character type — Sun et al. (2018b) partitioned their data according to whether a character described an evolutionary novelty or loss of a feature, or the transformation of a pre-existing feature (see electronic supplementary material (Sun et al., 2018a)). Moysiuk and Caron (2019) followed the same partitioning strategy, which is based on Sereno (2007) who proposed a fundamental distinction of two different types of morphological characters — neomorphic and transformational.

Homoplasy

Homoplastic characters are thought to show more random variation and thus carry less information about the evolutionary history of a clade (Farris, 1969). Limiting the weight given to them can increase the stability of phylogenetic inference in parsimony analyses (Goloboff et al., 2008). Rosa, Melo, and Barbeitos (2019) found that partitioning by homoplasy on a maximum parsimony tree outperformed alternatives in a Bayesian analysis.

Chapter 3

Methods

The aim of this thesis is to evaluate the performance of different partitioning strategies in phylogenetic inference. Secondarily I tested the effect of varying the branch length prior, which affects a large number of parameters in Bayesian phylogenetic models.

My approach centered on estimating the marginal likelihood of each combination of partitioning strategy and branch length prior using the techniques MC^3 and Steppingstones sampling in MrBayes. I also assessed relationships between model fit, branch lengths, and tree topologies, and investigated a connection between homoplasy and other measures of character structure.

3.1 Terms and definitions

3.1.1 Features of a phylogenetic tree

A phylogenetic tree is a graph expressing an evolutionary hypothesis. The taxa of interest sit at the tips of the tree. Branches represent the putative lineages connecting them to each other. Two branches diverge at an interior node, which represents the hypothetical last common ancestor (LCA) of its two descendant lineages. The root represents the LCA of all taxa on the tree.

3.1.2 Types of trees

3.1.2.1 Preferred tree

Each matrix was taken from a publication that contained a phylogenetic tree calculated from that matrix. This tree was designated the author's preferred tree and used as a starting point for several analyses. It performs a similar function as the true tree in a simulation study by providing a reference point to which all reconstructed trees can be compared.

3.1.2.2 Partitioning tree

The number of changes in a character on a tree contains information about its reliability for phylogenetic reconstruction. If it shows many more changes than the

minimum required, it is considered less reliable than characters that change state less often. A homoplasy index quantifies the amount of unnecessary change in a character, allowing us to rank and partition characters. To do this, we need to provide a tree topology on which to measure homoplasy. This tree is here called the partitioning tree. Each homoplasy-partitioned model uses a unique partitioning tree, ranging from the preferred tree for a dataset to a random tree, and trees generated by perturbing the preferred tree to various degrees (see Moves, section 3.1.2.4).

3.1.2.3 Reconstructed tree

A reconstructed tree here designates a tree recovered by Bayesian analysis. In the context of tree perturbations (see section 3.5 and chapter 4) this is the consensus tree found in a `.con.tre` file. The 50% majority consensus tree summarises the clades resolved in at least 50% of trees sampled by a MrBayes MC^3 analysis. By combining the most frequent clades and retaining polytomies where the topology is less reliably reconstructed, it balances precision and accuracy (Smith, 2019a) and represents a conservative estimate of the true topology.

3.1.2.4 Moves

Moves are rearrangements that transform one tree topology into another. I used the moves listed below to generate trees at various distances from the preferred tree topology:

Nearest Neighbour Interchange (NNI)
: A tree is divided into four subtrees connected by one interior branch. This branch and its connections are erased. The subtrees can then be reconnected in two alternative ways to create a neighbour tree. This perturbation move creates the least amount of topological change because the topology of the subtrees remains stable and only their relationship is changed.

Subtree Pruning and Regrafting (SPR)
: The tree is divided into two subtrees by breaking an interior branch. The broken branch of one tree is held onto. That subtree is grafted to the other by the broken branch to any branch. The topology of the first subtree remains unchanged but is moved into a new position on the second subtree.

Tree Bisection and Reconnection (TBR) The tree is divided into two subtrees by breaking an interior branch. One subtree is reconnected to the other by the middle of any branch to any branch. This move changes the topology more strongly than an SPR move because the topology of each subtree can be 'shuffled' before reconnection.

3.1.2.5 Comparing tree topologies

A phylogenetic tree can be viewed as a collection of four-taxon subtrees, or quartets. Similarity between two given trees can be expressed based on the numbers of unique and shared quartets. I calculated tree similarity using Quartet Similarity (QS), the complement of the Quartet Divergence (QD) as implemented in the `Quartet` R package (Smith, 2019d). The QD is calculated as Robinson & Foulds' distance metric applied to quartets (Robinson and Foulds (1981) developed the metric for bipartition splits), normalised against the number of quartets that would be resolved if two trees were identical:

$$RF = 2d + r1 + r2$$
$$QD = RF/2Q$$
$$QS = 1 - QD$$

where

d = number of quartets resolved in both trees

r1 = number of quartets resolved in tree 1 but not in tree 2

r2 = number of quartets resolved in tree 2 but not in tree 1

N = number of species

Q = total number of quartets = $\binom{N}{4}$.

3.1.2.6 Calculating tree length

The length of a tree is defined as the sum of all branch lengths. It can be calculated in R as the sum of the elements in the component `$edge.length` of a tree object of class "phylo".

3.1.3 Model fit

To compare phylogenetic models with different priors and parameters, we need a measure of how well they model real-world evolutionary processes. As we usually cannot observe the evolutionary history of a group of taxa directly, we rely on statistical measures of model fit.

3.1.3.1 Marginal Likelihood (ML)

A likelihood is a measure of the support that the data lend to a particular set of parameter values. The Marginal Likelihood of a Bayesian model is the likelihood of the data given that model, averaged over the whole parameter space (Fourment et al. (2019), and see section 2.3.2.1).

Each parameter addition adds a dimension to parameter space. If a model with 20 parameters estimates a phylogeny just as well as a model with 10 parameters, the more parameter-rich model receives a lower marginal likelihood because the likelihood is averaged over a larger space. Thus the marginal likelihood penalises over-parameterisation — the inclusion of parameters which do not markedly improve inference (Oaks et al., 2019).

Calculating the marginal likelihood directly involves a high-dimensional integral which is almost universally intractable in phylogenetic applications due to the number of parameters. It is usually estimated indirectly by sampling from the posterior distribution. Fourment et al. (2019) benchmark 19 methods of computing MLs on tree topologies. Here, the Steppingstones sampling method by Xie et al. (2011) was used. Marginal likelihoods in phylogenetics are very small and usually expressed in natural log units (logML).

3.1.3.2 Bayes Factor

To directly compare two models, the difference between their log Marginal Likelihoods is calculated. The resultant quantity, the ratio between their MLs, is the Bayes factor (BF). It quantifies the strength of evidence for one model over the other. Bayes factors have been confirmed to be a suitable method for choosing between competing partitioning strategies (Brown and Lemmon, 2007). Being a ratio of likelihoods, the Bayes factor is expressed in decibel (dB). Kass and Raftery (1995) provide guidelines for interpreting Bayes factors (see Table 3.1).

Every unit BF corresponds to a factor 10 difference in likelihood between two models. For example, a BF of 5 dB signifies that the fit of one model to the data is 10^5 times better than another.

TABLE 3.1: Interpreting Bayes factors according to Kass and Raftery (1995). They suggest taking double the natural logarithm of the BF. I have reported BFs as natural logarithm and have modified their suggested BF values.

2*BF	BF	Interpretation
$0-2$	$0-1$	Not worth more than a bare mention
$2-6$	$1-3$	Positive
$6-10$	$3-5$	Strong
>10	>5	Very strong

3.2 Bayesian analysis

Bayesian phylogenetic analyses using Metropolis-Coupled Monte Carlo Markov Chains (MCMCMC or MC^3) (see section 2.3.1.2) and Steppingstone sampling (Xie et al. (2011), and see section 2.3.2) were run in MrBayes v3.2.6 x64 (Ronquist et al., 2012). Examples of all types of files, including MrBayes files in NEXUS format (Maddison, Swofford, and Maddison, 1997), can be found online (see link in section 3.7.3)

3.2.1 Parameter settings in MrBayes

MrBayes receives data and instructions from files written in NEXUS format (Maddison, Swofford, and Maddison, 1997). They are made up of blocks variably containing character matrices, trees, parameters, and programmed instructions. The following subsection discusses the parameter settings used in the different types of analyses and references the MrBayes Manual throughout (Ronquist, Huelsenbeck, and Teslenko, 2011).

3.2.1.1 Ascertainment bias

Ascertainment bias or coding bias describes the bias introduced by coding only variable characters. Variable characters are those which are present in at least two states among the taxa of interest.

In molecular datasets, invariable characters are common and necessary for proper alignment of genetic sequences. When assembling morphological matrices on the other hand, often only characters showing variation within the surveyed group are included. Therefore the Mk model, being originally designed for molecular data, expects a proportion of invariable characters. MrBayes allows us to correct this bias by specifying whether a matrix contains any variable characters. When analysing discrete morphological character data, this is the default setting.

When only variable characters are coded, the underlying model is referred to as an Mkv model — an Mk model with only variable data.

```
datatype=standard;
lset coding=variable;
```

3.2.1.2 Rate variation

Morphological characters can vary considerably in their evolutionary rate of change. The evolutionary rate affects how likely they are to change state along a branch of a given length. On a short branch, we might expect to see a change of state in a character with a fast rate of change, but would assign a lower probability to seeing a change in a slowly evolving character. This diversity of evolutionary rates can be modelled through Among-character rate variation (ACRV).

I enforced gamma-shaped ACRV. This lets the rate of change of each character vary under a Γ-distribution with a single shape parameter α. MrBayes optimises ACRV by sampling values for the shape parameter α from a flat prior over the interval (0.5, 50).

```
lset rates=gamma;
```

In partitioned analyses, I additionally modeled Among-partition rate variation (APRV). ACRV is estimated for each individual partition. Rate multipliers link partitions so that the average rate of change among partitions remains 1. By default, the rate multipliers are drawn from a flat Dirichlet prior where all values are 1.0. APRV is accommodated by setting:

```
prset applyto=(all) ratepr=variable;
```

In partitioned models, I allowed ACRV to vary from partition to partition by unlinking the shape parameter α. Rates of change and therefore branch lengths are drawn from a unique Γ-distribution for each partition, a per-partition Γ.

```
unlink shape=(all);
```

3.2.1.3 Branch lengths

Rannala, Zhu, and Yang (2012) implicate misspecified branch length priors in causing convergence problems in the tail of the posterior distribution. As branch lengths increase, the posterior flattens out and the likelihood approaches a constant. Chains may become stuck in this flat "tail" of the posterior, thus leading to insufficient convergence. Ekman and Blaalid (2011) also argue for the importance of the branch length prior in affecting clade support as well as posterior probabilities. They advocated for the use of gamma dirichlet priors favouring short branch lengths, which for their data resulted in trees with strongly supported clades and short, consistent tree topologies.

A meaningful branch length prior should reflect the expected mean number of character changes per branch. We can obtain an estimate of this value by dividing the parsimony score of the most parsimonious tree for a data set by the number of characters. The prior is then parameterised with the inverse of this value (Sun et al., 2018a).

$$\lambda = \left(\frac{\text{Length of Maximum Parsimony tree}}{\text{number of characters}}\right)^{-1}$$

To evaluate the effects of partitioning under different branch length priors, I tested four different model specifications:

Default — The default branch length prior in MrBayes v3.2.6 is a gamma dirichlet distribution with shape parameter $\alpha_T = 1$.

Fitted Gamma — I tested a variation of this model where α_T is fitted to the data set based on the number of characters and the length of a Maximum Parsimony tree (λ, see equation 3.2.1.3).

```
prset applyto=(all)
brlenspr=unconstrained:gammadir(1,λ,1,1)
```

Fitted Exponential I also employed an exponential distribution parameterised by λ.

```
prset applyto=(all) brlenspr=unconstrained:exp(λ);
```

Relaxed Clock I tested a morphological clock model as introduced by Lee et al. (2014). I followed the parameter settings of those authors in using a relaxed clock model with independent gamma rates (IGR). Rate variation across lineages is independent, but the rate at which the expected variance of the branch lengths increases over time is restricted by the parameter `Igrvarpr`. The prior on `Igrvarpr` was set to an exponential distribution with rate λ. The prior for the clock rate was set as a normal distribution with a mean of 0.01 and a standard deviation of 0.1. This distribution is quite wide and flat, so only weakly constrains the posterior.

```
PRSET applytp=(all) brlenspr=clock:uniform;

PRSET applytp=(all) clockvarpr=igr;
PRSET applytp=(all) igrvarpr=exp(λ);
PRSET applytp=(all) clockratepr=normal(0.01,0.1);
```

TABLE 3.2: All datasets were tested under the default gamma dirichlet prior with shape parameter $\beta_T = 0.1$. I also tested a gamma dirichlet prior using λ as the shape parameter β_T, an exponential prior with rate λ, and a relaxed clock model with independent gamma rates (IGR).

Branch Length Model	**HYO**	**CEA**	**OZL**	**SCO**	**THER**
IGR relaxed clock	X	X	X		X
unconstrained:gammadir(1,0.1,1,1) (default)	X	X	X	X	X
unconstrained:gammadir(1,λ,1,1)			X	X	X
unconstrained:exp(λ)	X	X		X	

3.2.1.4 Topology

The `outgroup` command lets us specify a root by providing the outgroup taxon.

```
outgroup Plumariidae;
```

Topology priors fragment the posterior because they severely depress the likelihood of certain clade groupings that do not conform to the constraint. This rugged posterior can make it more difficult to achieve proper mixing and convergence. It

is possible to set a prior on topology by enforcing the grouping of certain taxa, as was done for dataset OZL following Lee et al. (2014):

```
constraint root = 1-.  ;
constraint birds = 2-.  ;
constraint Avebrevicauda = 3-.  ; [ingroup]
PRSET topologypr = constraints(root, birds, Avebrevicauda);
```

Ages of terminals for OZL were calibrated to their earliest stratigraphic ages (see Lee et al. (2014) Supplementary Information 1).

```
CALIBRATE Dromaeosauridae = fixed(111);
CALIBRATE Archaeopteryx = fixed(150);
CALIBRATE Ambiortus = fixed(130);
...
CALIBRATE Zhongornis = fixed(122);
```

The age of the root divergence was calibrated using a uniform prior spanning 150.1–200 Mya.

```
CALIBRATE root=uniform(150.1,200);
```

Finally, I enforced the calibrations of all node ages.

```
PRSET nodeagepr = calibrated;
```

3.2.1.5 State frequencies

In the overarching model used by MrBayes for morphological data, state frequencies are assumed to be equal. Thus for a binary character, the model would assume an equal probability for a state to be 0 or 1. In the case of a neomorphic partition where one state is the ancestral condition, this may not be a proper prior.

MrBayes allows us to specify an unequal state frequency prior using a symmetric Dirichlet distribution. It takes a single shape parameter α_{symdir} which specifies the degree of asymmetry across sites. I set α_{symdir} to 1.0 for the neomorphic partition, which applies a uniform prior across the different state frequency proportions.

```
prset applyto=(neomorphic partition) symdirihyperpr=fixed(1.0);
```

Relaxing the equal state frequency constraint adds a parameter. The model is penalised unless the addition improves inference.

3.2.1.6 Ordered characters

Several matrices contained characters with specified state orders. In unordered characters, the transition rate is the same no matter which state changes into which

other. A transition matrix for an unordered character with three states (after Lewis (2001)):

$$Q_{unordered} = \alpha \begin{bmatrix} - & 1 & 1 \\ 1 & - & 1 \\ 1 & 1 & - \end{bmatrix}$$

For ordered characters, transitions are only possible between adjacent states:

$$Q_{ordered} = \alpha \begin{bmatrix} - & 1 & 0 \\ 1 & - & 1 \\ 0 & 1 & - \end{bmatrix}$$

To effect a change from state `1` to state `3`, the lineage must traverse state `2`. Thus a transition from `1` to `3` requires one extra step.

3.2.1.7 Parameters for MC^3

The number of generations was set to 5,000,000 with a tree sampled every 200 generations. 8 parallel runs were executed and each run contained 8 metropolis-coupled Markov chains. The first 25% of generated trees were discarded as burnin and not considered during parameter estimation.

```
mcmcp ngen=5000000 samplefreq=200 nruns=8 nchains=8 burninfrac=0.25
append=no temp=0.05;
mcmc;
```

Summary files of all tree and parameter estimates were requested.

```
sumt;
sump;
```

3.2.1.8 Parameters for Steppingstones sampling (Xie et al., 2011)

The number of generations was set to 5,000,000 with a tree sampled every 100 generations. 8 parallel runs with 8 chains each were used. The first 25% of generated trees were discarded as burnin.

```
mcmcp ngen=5000000 samplefreq=100 nruns=8 nchains=8 burninfrac=0.25
append=no;
```

Tree space was sampled moving from the prior towards the posterior, gradually providing the model with more data. I chose to do so because convergence problems arose when moving from the posterior towards the prior as is the default in MrBayes. These problems are discussed below in section 3.2.2.

```
ss burninss=-1 nsteps=40 fromprior=yes;
```

All other model parameters were held constant across data sets and partitioning strategies, except where chain and run behaviour induced us to lower the temperature parameter to improve convergence (see section 3.2.2 below).

3.2.2 Convergence

When running Steppingstone analyses from posterior to prior, between 50–75% of runs returned positive log marginal likelihoods. These equate to likelihoods much larger than 1. Likelihoods greater than 1 are technically possible, but these values were much higher than any of the marginal likelihoods estimated by MC^3. MC^3 tends to estimate higher marginal likelihoods than Steppingstone sampling and thus served as a useful reference point against which to check these inflated results. Runs with high likelihoods usually showed poor chain mixing behaviour, indicating that these runs hadn't converged properly. It appears most likely that all of the run's chains were stuck in an isolated area of the posterior distribution for several steps. As the analysis moved one stepping stone closer to the prior and found a flatter power posterior, the chains could now move out of their region and discover the rest of the distribution. Recall that in steppingstone sampling the marginal model likelihood is calculated as the product of the ratios of power marginal likelihoods of subsequent steps (see equation 2.7). Once the previously isolated run can access the entire power posterior, it calculates a much larger power marginal likelihood than on the previous step, resulting in a ratio considerably greater than 1. To ensure that runs discover the entire power posterior from the beginning, the final Steppingstone analyses were run from the prior to the posterior, which significantly reduced the number of positive logMLs encountered. Any remaining positive runs were discarded before calculating mean marginal likelihoods.

I monitored acceptance rates for moves in the cold chain of each run. Acceptance rates were generally within acceptable limits of 20–60% (Ronquist, Huelsenbeck, and Teslenko, 2011). In addition, I checked chain swap rates to ensure proper mixing of the individual chains within each run. In several instances where chain swap rates were lower than 0.1, the temperature to which chains are heated was decreased to `temp=0.05` from the default of `0.10`. This increased the proportion of accepted chain swaps and improved mixing of the cold chain.

I used the average standard deviation of split frequencies (split) printed to `.ss` and `.out` files to diagnose whether independent runs had converged. Values below 0.01 are generally accepted as a sign of very good convergence, and values up to 0.05 are acceptable (Ronquist, Huelsenbeck, and Teslenko, 2011). In some Steppingstone analyses, split values exceeded these limits. To test whether the analyses had nevertheless converged sufficiently, I reran several analyses with high split values under two schemes with changed settings. Under one scheme, models had a higher burnin proportion (HB), while models under the second scheme were run for more generations (MG). Under the HB scheme, I increased the burnin proportion at the beginning of each step from 25% to 50%. The number of generations was also increased so that the same number of samples was collected during each step as before. Under the MG scheme, analyses were run for 10,000,000 generations rather than 5,000,000 as before. The script `ConvergenceTesting.R` was used to extract split values and calculate marginal likelihoods from the result files.

Both types of models showed improved convergence (see Table B.1), but the marginal likelihoods of most MG and HB models did not differ strongly from those of the standard models. This strongly suggests that the standard settings of a burnin proportion of 25% and 5,000,000 generations were sufficient for most models to converge. Furthermore, results should be comparable at least among analyses of the same dataset and branch length prior since split values were quite similar within these groupings.

3.3 Datasets

I chose five morphological character matrices that were published in the context of a phylogenetic analysis. The studies had different foci. Some investigated certain evolutionary hypotheses, such as mosaic evolution (Clarke and Middleton, 2008) or morphological clocks (Lee et al., 2014), while others simply had the aim of calculating a phylogeny.

The aim of this thesis was to provide simple rules for partitioning morphological data that may be applied to any dataset, regardless of its structure or character selection strategy, to improve phylogenetic inference. Numbers of taxa ranged from 16 to 65 and numbers of characters from 27 to 242. Most datasets included a high proportion of fossil taxa as these are most commonly analysed using morphological data.

3.3.1 Character coding

Flaws in character selection and formulation were present in several of the datasets used. Compound characters posed the most common character coding problem. These characters collapse character states which share a hierarchical relationship into one character (Brazeau, 2011).

Example of a compound character (from Yates (2007)):

TABLE 3.3: The datasets included in this study. Column "NT" gives the number of taxa, "NC" the number of characters. The ratio of taxa to characters is a proxy for a dataset's signal strength and is given in column "Signal".

Abbr.	Group	NT	NC	Signal	Reference
HYO	Lophotrochozoans	56	225	0.249	Sun et al. (2018b)
CEA	Birds	25	205	0.122	Clarke and Middleton (2008)
OZL	Birds	65	247	0.263	Lee et al. (2014)
SCO	Scolebythid Wasps	16	27	0.593	Engel, Ortega-Blanco, and McKellar (2013)
THER	Theridiid Spiders	64	242	0.264	Agnarsson (2006)

> Splenial foramen: *absent (0), present and enclosed (1), or present and open anteriorly (2)*.

For a human reader, this formulation makes intuitive sense, but in computational analyses of morphological data, compound characters pose a problem because the grouping of the two 'present' states is lost (Brazeau, 2011). Applying a 3-state model to the above character assumes equal probabilities of transitioning from state $0 \Longleftrightarrow 1$, from $1 \Longleftrightarrow 2$, and from $0 \Longleftrightarrow 2$. But is losing an anteriorly open splenial foramen really as likely as transitioning from an enclosed foramen to an open one?

An improved formulation of the splenial foramen character separates it into a neomorphic and a transformational character:

> Splenial foramen: *absent (0), present (1)*. (neomorphic)
> Splenial foramen, degree of enclosure: *enclosed (0), open anteriorly (1), inapplicable (-)*. (transformational)

Brazeau (2011) provides a comprehensive review of character formulation problems, their impacts on phylogenetic inference, and advice on improving character statements. The transformational character is inapplicable where the splenial foramen is absent. For recommendations on handling inapplicable characters in phylogenetic analysis of morphology, refer to Brazeau, Guillerme, and Smith (2019).

For this analysis, most character coding flaws were not corrected due to time constraints.

3.3.2 HYO (Sun et al., 2018b)

Sun et al. (2018b) assembled a character matrix to illuminate the placement of a novel orthothecid hyolith fossil. The dataset broadly samples Lophotrochozoan taxa, including molluscs, brachiopods, phoronids, and sipunculans. The authors followed Brazeau, Guillerme, and Smith (2019) in using reductive coding (*sensu* Wilkinson (1995)), where inapplicability of a character is coded via a distinct state. Where a neomorphic character was inapplicable as a result of ontological dependence (*sensu* Vogt (2018)), this was coded as absence.

I deviated from the coding procedure of Sun et al. (2018b) in coding neomorphic as well as transformational character states starting from 0. When modelling the evolution of discrete morphological characters, MrBayes determines the number of states by finding the largest state code. If a character's states are not coded starting from 0, MrBayes will wrongly choose the model for a character with one more state than appropriate for that character.

The matrix contains 54 taxa which were coded for 225 morphological characters.

The shape parameter for the branch length prior was calculated to be $\lambda = 0.3$. All characters were unordered. The preferred tree was drawn after Figure 4 of Sun et al. (2018b), the consensus of trees constructed using Maximum Parsimony. Bayesian analysis recovered a compatible tree topology.

3.3.3 CEA (Clarke and Middleton, 2008)

This bird dataset was assembled by Clarke and Norell (2002) and extended by Clarke, Zhou, and Zhang (2006). Clarke and Middleton (2008) investigated the role of anatomical mosaicism in Avian evolution by reanalysing the matrix using anatomical partitioning.

25 taxa were scored for 205 characters. 20 of the 25 included taxa were scored based on fossil evidence. The 205 characters include 38 ordered characters (see Appendix A.1.2).

The authors of the matrix did not distinguish between missing data and inapplicable characters. The fitted branch length priors were parameterised with $\lambda = 0.5$ (see equation 3.2.1.3). The preferred tree is in Figure 3b of Clarke and Middleton (2008). It was inferred by anatomically partitioned Bayesian analysis.

3.3.4 OZL (Lee et al., 2014)

This second bird dataset is unique among my selection of matrices in that it uses earliest fossil appearances to date fossil terminals and infer ages for interior nodes of the tree. They used a relaxed morphological clock in MrBayes to constrain the timing of the radiation of crown Aves.

65 taxa were scored for 247 characters. 22 of them were also included in the dataset CEA. 60 out of 65 taxa were fossils.

31 characters were ordered (see section A.1.3). O'Connor and Zhou (2013) ordered 32 characters which included character 64. However, this character has more than six states and cannot be ordered in MrBayes. Thus it was treated as unordered here. No distinction was made between missing data and inapplicable characters.

I enforced a topology prior following the original authors. Lee et al. (2014) specified the topology at the root, outgrouping Dromeosauridae, placing Archaeopteryx at the root of the bird clade, and constraining the remaining taxa to form the ingroup. The minimum ages of terminal taxa were calibrated using first appearances in the fossil record. For details on node dating, refer to section 3.2.1.4. The nexus files contain more detailed information and are found online.

The branch length prior was fitted to the dataset with the shape parameter $\lambda = 0.3$. The preferred tree was drawn after Figure 2 of Lee et al. (2014), which was inferred using Bayesian analysis.

3.3.5 SCO (Engel, Ortega-Blanco, and McKellar, 2013)

Engel, Ortega-Blanco, and McKellar (2013) focused on reconstructing the relationships within a comparatively small family of wasps. Scolebythiidae comprises just five recent species, but a number of fossils in amber from the Early Cretaceous and Eocene have been assigned to it as well. The matrix used here is an extension of a dataset assembled by Engel and Grimaldi (2007).

10 out of 16 taxa are known only from the fossil record. Taxa were scored for 27 characters. The coding of characters 9 and 21 for the outgroup Plumariidae was corrected from (0123) to (01) as these two characters only have two states. All characters were left unordered. Character inapplicability was treated like missing data.

The branch length prior was parameterised with $\lambda = 0.4$. The preferred tree is shown in Figure 6 in Engel, Ortega-Blanco, and McKellar (2013). It is the strict consensus between three most parsimonious trees.

3.3.6 THER (Agnarsson, 2006)

This study placed a novel genus into the wider phylogeny of Theriidid spiders. It extended the character matrix of Agnarsson (2004) by 2 species. This matrix is the only one studied here which does not include any fossil taxa, and is further unique for routinely including several species from the same genus. It is the most fine-grained matrix of all those covered here, setting a counterpoint to the hyolith matrix of Sun et al. (2018b) which samples broadly from the Lophotrochozoan clade including brachiopods, molluscs, and worms.

64 extant taxa were scored for 242 characters. There were no unordered characters. The authors distinguished between inapplicable characters and missing data (reductive coding *sensu* Wilkinson (1995)).

The shape parameter of the branch length prior was $\lambda = 0.3$. Figure 11 in Agnarsson (2006) shows the Maximum Parsimony tree which was used as the preferred tree.

3.4 Partitioning Strategies

Table 3.4 lists the types partitioning strategies that were tested on the five morphological character matrices described above.

TABLE 3.4: Partitioning strategies tested using Steppingstones Analysis Xie et al. (2011).

Partitioning Strategy	Partitions	Example of usage
Homoplasy	3 — Low, mid, high f_i	Rosa, Melo, and Barbeitos (2019)
Character type	2 — Neomorphic, Transformational	Sereno (2007)
Anatomy	3–8 — e.g. head, trunk, limbs	Clarke and Middleton (2008)
Information Content	3 — low, mid, high IC	—
random	3	—

3.4.1 Unpartitioned analysis

Each matrix was first analysed without partitioning the characters. The result of the unpartitioned analysis served as the baseline against which partitioned models were compared (see section 3.1.3.2).

3.4.2 Homoplasy partitioning

Under this partitioning scheme, characters were ranked by their homoplasy on a given tree. Homoplasy indices quantify the amount of "unnecessary" evolutionary change a character goes through. The more times the same character state evolves independently or reverts to the ancestral state, the more homoplastic a character is. A character with no homoplasy shows exactly as many changes of state as necessary given a tree topology and the terminal character states. If a character changes state more often than required, it is deemed homoplastic and considered to evolve at a higher rate.

The homoplasy index used here is Goloboff's unbiased measure of homoplasy, f_i for a character i (Goloboff, 1993):

$$f_i = \frac{k+1}{s_i + k + 1 - m_i}$$

where

k = concavity constant

s = observed number of changes of state, or steps, on a tree

m = minimum number of steps required on a tree.

The concavity constant k was set to 5. Observed steps (s) and minimum steps (m) are computed using the Fitch algorithm (function `Fitch()`) and the function `MinimumSteps()` in the R package `TreeSearch`. The Fitch algorithm for parsimony allows for reversals and assumes symmetric change among character states (Wagner Parsimony, see Felsenstein (2001)). It calculates the number of changes of state (steps) in each character on a tree via post-order tree traversal. Node numbers refer to Figure 3.1.

1. We assign to each tip the set of character states observed in that taxon. Node 1 shows state A, so its set is {A}. Node 2 is polymorphic and shows both states A and B, so the set is {AB}.

2. We move rootwards to the first interior nodes and compare the sets of the two descendent tips. These parent nodes are assigned the intersection of the descendent sets. The descendent sets of node 5 are {A} and {AB}, therefore the parent node is assigned set {A}. If the intersection is empty, the parent node is assigned the union of the descendent sets and we count one change

of state. Node 6 is assigned the set {AC} since its descendents 3 and 4 have sets {C} and {A}.

3. We traverse the entire tree in this manner, counting a change of state every time we encounter an empty intersection and assign the union of the descendent sets to the parent node.

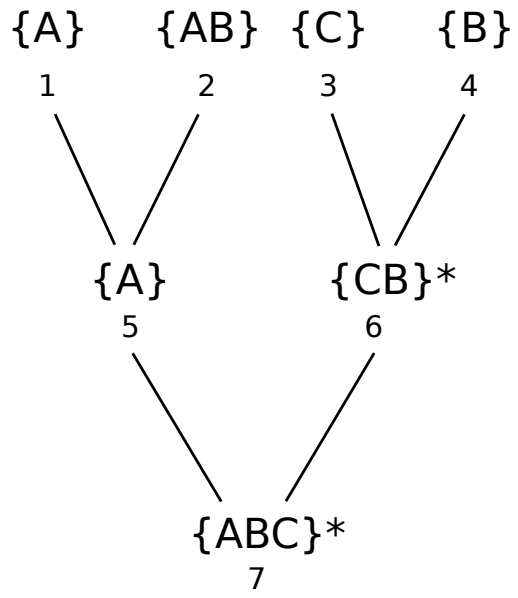


FIGURE 3.1: The Fitch algorithm for counting steps in a character on a topology by post-order tree traversal. A step (marked with *) is counted at every node whose two descendent nodes do not share a character state.

Minimum steps on a given tree can be calculated based on the number of states in a character. If a character has two states, at least one change of state must occur, from the ancestral to the derived state. For a character with 5 observed states, there must be a minimum of 4 changes. Thus, where j = the number of states of a character the formula for minimum steps is simply

$$m = j - 1$$

Goloboff's f is favoured here over Farris' consistency index (CI) (Farris, 1969). When calculating the CI, extra changes in binary characters carry more weight compared to extra changes in multistate characters. In contrast, f treats binary and multistate characters identically and confers no more or less weight on changes of state in either type.

To facilitate comparison among characters, I normalised each character's homoplasy index against the maximum possible number of steps. The number of states in a character and the number of taxa put a limit on the number of steps that can be attributed to homoplasy, thus placing an upper bound on a homoplasy index (Hoyal Cuthill, Braddy, and Donoghue, 2010):

> "The highest number of steps possible for a character is then $g = t - F$, where F is the number of taxa with any one most frequent state (Steel and Penny, 2006). The total number of taxa with a minority state will be maximal for the most even distribution of states possible. Under this most even state distribution, $t - F$ is equal to the maximum possible number of steps that the character may show on any tree (g_{max}).
>
> $$g_{max}(t;n) = t - [t/n]$$
>
> where
>
> $$[t/n] = \text{ the smallest integer} \geq t/n$$
>
> [$[t/n]$ is] equal to the lowest possible number of taxa with any one most frequent state (F_{min})."
>
> — from Hoyal Cuthill, Braddy, and Donoghue (2010)

For each character, the observed homoplasy index f_{obs} was calculated using the observed number of steps s_i and the minimum number of steps m_i. Using the maximum number of steps g_{max}, the upper bound of the homoplasy index f_{max} was calculated. f_{obs} was divided by f_{max} to arrive at a measure of homoplasy normalised to the number of states and taxa:

$$f_i = f_{obs}/f_{max}$$

3.4.2.1 ... based on preferred tree

I calculated f_i for each character on the preferred tree (see section 3.1.2.1). The characters were ranked from high to low homoplasy index and sorted into three partitions of equal size using the script `Workflow.R`.

3.4.2.2 ... based on random tree sample

Partitioning by homoplasy on single tree might bias the model towards similar trees. Furthermore, it seems circular to use a tree to estimate a tree. A group of random trees could contain the same information on character reliability without potentially favouring a certain phylogeny and distorting the posterior sample. To investigate whether the success of homoplasy partitioning (Rosa, Melo, and Barbeitos, 2019) is caused by this circularity, I tested a model partitioned by homoplasy on a sample of random trees.

Random trees were generated by recursive random splitting of edges using the function `rtree()` in the R package `ape`. I calculated f_i on 300 random trees and the took the arithmetic mean F_i. All characters were ranked by their F_i and divided up into three equal partitions with the script `fPartitioning.R`.

3.4.3 Neomorphic-Transformational partitioning

Sereno (2007) posits that there are two fundamentally different patterns of morphological evolutionary change — the arisal of novel structures, and the modification of existing features. This distinction is reflected in the structure of character statements. A neomorphic character describes whether a structure is present or not — the *de novo* appearance or loss of a morphological trait. Such characters only have the states *absent* and *present*. The numeric code 0 is assigned to that state which is inferred to be the ancestral condition.

In contrast, a transformational character describes a change in an existing structure. This change may be quantitative, for example a change in size, or qualitative, such as a change in shape, appearance, or composition.

Any morphological character can be stated according to the schema laid out by Sereno (2007), allowing the division of any data set into two partitions. These contrasting character types are thought to evolve at different rates: it is easier to change the shape of a vertebra than to evolve a spinal column.

3.4.3.1 . . . with asymmetric transition rates

Neomorphic characters might be expected to change preferentially from the ancestral to the derived condition while reversals are rare (see for example Sun et al. (2018a)). This inequality can be modelled by setting an asymmetric hyperprior on the transition rate for the neomorphic partition (see section 3.2.1.5).

3.4.3.2 . . . with symmetric transition rates

On the other hand, reversals may be equally common as transitions from ancestral to derived states. I also tested a version of neomorphic-transformational partitioning where both partitions were modelled under a symmetric transition rate prior.

3.4.4 Anatomical partitioning

Anatomy-based partitioning can be employed to better model mosaic evolution where change is localised to a region of the body and related characters evolve in concert (Felice and Goswami, 2018). This concept is related to morphological integration, the functional linkage of characters where a change in one character accompanies or leads to a change in another. Such characters would have similar evolutionary rates (Miller and Olson, 1960). Partitioning by anatomy has been used in several phylogenetic studies using Bayesian methods, for example Gaubert et al. (2005) and Clarke and Middleton (2008).

Characters were sorted into partitions manually. After gaining an overview of the dataset and different characters, I attempted to sort them into as few partitions as was possible while maintaining sensible anatomical divisions. This was more or less straightforward depending on the character selection criteria used by the

authors of the original publications my datasets were drawn from. Characters describing ratios or relationships between features were allocated to the partition of the principal structure. The dataset THER included behavioural characters, which were allocated to a partition of their own under this scheme. I also tried to group characters in a way that would leave no partitions with fewer than 6 characters so that parameters were estimated on the basis of enough data. The list of character partitions for all datasets can be found in Appendix A.

The number of partitions under this scheme is variable and highly dependent on character selection criteria. In this study, it ranged from 3 for SCO to 8 for HYO. Because the number of partitions is often arbitrary and higher than for other partitioning strategies, I also tested the two datasets with the most anatomical partitions with fewer partitions (OZL with 6 instead of 7 and HYO with 4 instead of 8 partitions).

3.4.5 Partitioning by Information Content

The Information Content (IC) of a character is measured as the number of tree topologies which are perfectly compatible with that character. A character is said to be compatible if it shows no unnecessary changes on a tree, that is, if the number of observed steps equals the minimum required number of steps $j - 1$ (Felsenstein, 2001). The reader will recall that any steps in addition to the minimum steps are homoplastic steps.

A character with k states is compatible with a tree if it shows $j - 1$ changes of state on that tree. A character that is compatible with only a small number of tree topologies restricts the possible tree topology more than a broadly compatible character. A character's information content can then be expressed as the proportion of compatible trees to all possible tree topologies (Smith, 2019. Personal communication). This measure is similar to Faith and Trueman (2001)'s central quantity of Profile Parsimony, where a character's number of steps on a tree are compared to the number of steps on a random tree.

I calculated each character's IC using the function `CharacterInformation()` in the `TreeSearch` R package (Smith, 2019c). The function calculates the difference between the number of unrooted trees compatible with a character and the total number of unrooted trees with N terminals where N is the number of taxa. I ordered characters by their IC and divided them into three partitions of equal size using the script `InformationContent`.

3.4.6 Random partitioning

Characters were put into a random sequence and divided into three equally sized partitions with the script `randomPartitions.R`.

3.5 Tree perturbation experiments

This experiment tested how the topology of the partitioning tree influences the performance of homoplasy partitioning. I used Metropolis-Coupled Monte Carlo Markov Chains (see section 2.3.1.2) to collect a large sample of trees from the posterior distribution. The consensus tree of this sample was designated the reconstructed tree (see section 3.1.2.3). I tested three datasets — HYO, SCO, and THER — and generated 400 partitioning trees for each, totalling 1200 Bayesian analyses. CEA and OZL were not tested here due to time constraints.

3.5.1 Tree generation and perturbation

The preferred tree for a dataset was coded into NEXUS format if there was no tree file provided in the supplementary information or uploaded on TreeBase.

Starting from this tree, perturbed trees were generated using different rearrangement moves (see section 3.1.2.4) implemented in the `TreeSearch` R package (Smith, 2019c). 100 random bifurcating trees were generated using the function `rtree()` from the R package `ape`. `rtree()` generates trees by random recursive splitting of edges (Paradis et al., 2019). The sample of trees is drawn from a uniform distribution. The perturbed trees also included three chains of 100 trees perturbed with NNI, SPR, and TBR moves respectively.

To create a chain of trees, the partitioning tree is perturbed once to create the first tree in the chain. This tree is then rearranged again using the same move in a new random location to generate the second tree in the chain. This is repeated until the chain contains 100 trees.

The R script `perturb_trees.R` contains the code to generate the different types of partitioning trees.

3.5.2 MC^3 analysis

Each perturbed tree served as the partitioning tree for a homoplasy-partitioned Bayesian analysis using MC^3. Characters were ordered by f_i as described in section 3.4.2 and allocated to four partitions. A NEXUS template specific to each data set was modified using the script `workflow.R` to implement these partitions in the MrBayes block. The parameter settings for MC^3 are detailed in section 3.2.1.7.

Partitioned MC3 analyses were run using the parallel version of MrBayes (Altekar et al., 2004) on Hamilton, Durham's High Performance Computing cluster.

3.5.3 Comparing model fit

Each MC^3 analysis generated several output files, among them a `.lstat` file to which the estimated log marginal likelihoods for all parallel runs are printed, and a `.con.tre` file which contains the consensus tree in NEXUS format.

To compare model fit, the arithmetic mean of the logMLs, which is equivalent to the harmonic mean of the MLs, was calculated for each analysis. The partitioning tree files (located in the subfolder `SingleStartTrees` within each dataset folder) and the `.con.tre` files were used to calculate Quartet Similarity between these and the preferred tree (see section 3.1.2.5). Model fit, tree similarity, and tree length were calculated and plotted using the script `TreePerturbationsResults.R`. The script `Correlation_TreeSim.R` checks whether the topologies of partitioning trees, preferred trees, and reconstructed trees are correlated.

3.6 Character structure

Character Information Content (see section 3.4.5) was used as a tree-independent way to quantify the amount of phylogenetic information contained within each character's state distribution over all taxa. The relationship with homoplasy was plotted with the script `CharacterStructure.R`.

3.7 Testing model fit under different partitioning strategies

I calculated model fit for a range of partitioned models and compared their performance to an unpartitioned model. Each dataset was partitioned according to eight strategies (including no partitioning) which are described in detail in section 3.4.

I used Steppingstone sampling (Xie et al., 2011) to estimate marginal likelihoods as accurately as possible. During preliminary experiments, MLs generated using the Steppingstones method were found to be much more accurate than the MLs calculated as part of a standard MC^3 analysis (see also Fourment et al. (2019)).

3.7.1 File preparation

A template specific to each dataset was modified to contain a block of code implementing partitions. For the neomorphic-transformational and anatomical partitioning strategies, characters were partitioned manually in an excel spreadsheet. The vectors of the character numbers in each partition were then copied into the template. For the other partitioning strategies, the template was modified using an R script (see each subsection within 3.4).

The MrBayes blocks of the files contained the parameters for Steppingstone sampling as detailed in section 3.2.1.8. Steppingstone analyses were run on the parallelised version of MrBayes on the Hamilton cluster.

3.7.2 Comparing model fit

Results were analysed and plotted using the scripts `partstratsCompare.R` and `partstratsCompareDefault.R`. For each partitioned analysis, the Bayes factor relative to the unpartitioned analysis of that dataset was calculated. The arithmetic

means of the ln marginal likelihoods, Bayes factors, and associated statistics were written to a summary file for each dataset (e.g. "HYO_MLsBFs.csv") with the script `partstratsCompare.R`. Pairwise comparisons of the average ranks of all partitioning strategies were plotted with the script `ranksHeatmap.R`. Model fit for each partitioning strategy was compared across datasets and branch length priors using the script `brlensCompare.R`.

3.7.3 Scripts

All relevant scripts, NEXUS templates, and summary filed for each dataset can be downloaded from GitHub at https://github.com/smf541/ThesisSupplements.

Chapter 4

Results I — Tree perturbations and character structure

4.1 Results of tree perturbations

Partitioning by homoplasy on a parsimony tree was named the best partitioning strategy in a comparative study by Rosa, Melo, and Barbeitos (2019). This type of partitioning scheme always relies on a partitioning tree (see section 3.1.2.2) on which the homoplasy score of every character is calculated. But what if that tree is unreliable? The chosen partitioning tree could be quite dissimilar from the true tree. How does this affect the tree inferred by Bayesian analysis?

I analysed three character matrices under homoplasy partitioning based on partitioning trees progressively further removed from the preferred tree (see section 3.1.2.1). Results from 1200 analyses across the three datasets are summarised below.

4.1.1 Evolution of the partitioning tree

Repeatedly perturbing the preferred tree for a dataset produces chains of more and more dissimilar partitioning trees. These trees range from identical to the preferred tree (Quartet Similarity = 100%) to effectively random (QS $\approx$ 33%). Any two random trees are expected to share a third of quartets because a quartet of any four taxa can be resolved in three equally likely ways (Smith, 2019b).

Depending on the perturbation move used (see section 3.1.2.4), a chain of perturbed trees reaches effective randomness at different speeds. NNI moves only make minor changes to the tree, so the 100th NNI tree for the HYO dataset still shares 80% of quartets with the preferred tree. 60 sequential SPR moves perturb the same preferred tree into an effectively random tree, and using TBR moves, it only takes 40 moves. The same holds true for the large THER tree. The SCO preferred tree is much smaller and reaches randomness within a handful of moves of any type.

4.1.2 Model fit is dependent on partitioning tree similarity

My results demonstrate that the fit of a homoplasy-partitioned model worsens as the partitioning tree grows more dissimilar to the preferred tree.

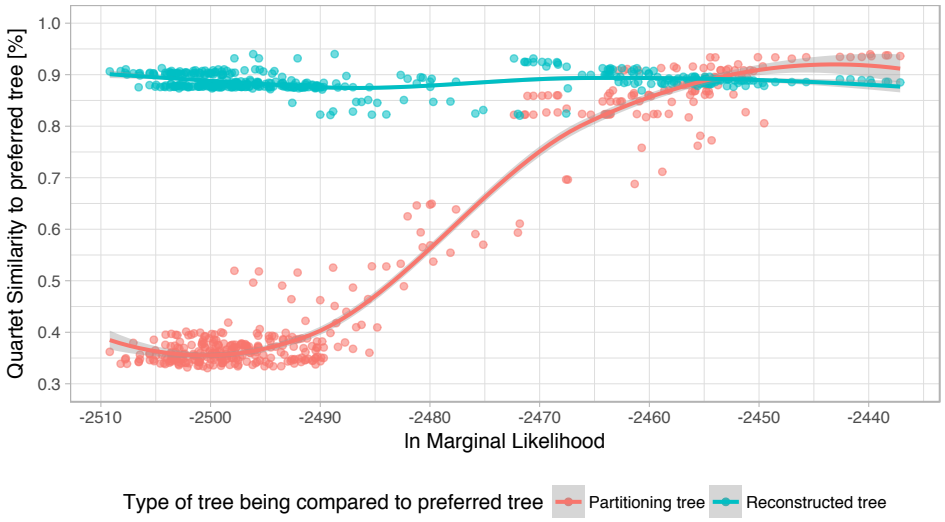
Fig. 4.1 shows the relationship between model fit and tree similarity (see 3.1.2.5) between the partitioning tree and the preferred tree as well as between the reconstructed tree and the preferred tree. The Bayes factor between the best-fitting and worst-fitting models is in the region of 70 dB for HYO, 10 dB for SCO, and 80 dB for THER. The models with poorest model fit are consistently based on partitioning trees that are quite dissimilar to the preferred tree — dissimilar enough to be effectively random.

Even models based on completely random trees still show considerable variation in model fit. For dataset THER, the fits of models based on random trees vary within a range of 30 dB though their partitioning and reconstructed trees do not vary in similarity to the preferred tree (see 4.1c). According to Kass and Raftery (1995), a Bayes factor of 30 dB suggests unequivocally to favour one model over another — but these models were created according to the same rules based on superficially identical trees. The question is then, what makes one model based on a random tree better than another?

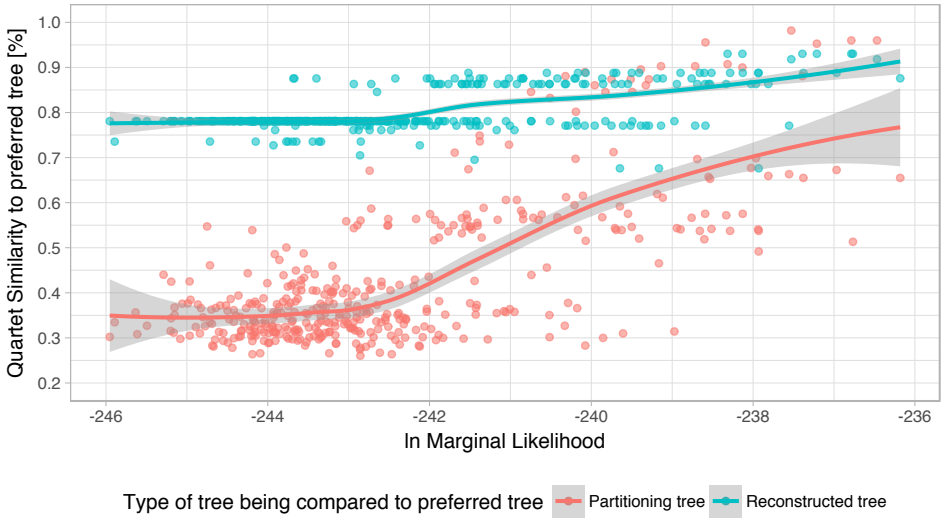
4.1.2.1 Interpretation

Model fit reflects whether a partitioning strategy successfully discriminates characters by their evolutionary rates, or whether they are sorted more or less randomly. If successful, partitioning aids inference by modelling variation in evolutionary rates, and the correct tree is arrived at readily. If character partitions do not reflect real rate categories, separately estimating the rate distribution for each partition is an unnecessary computational burden. This is penalised during marginal likelihood calculation, leading to worse model fit.

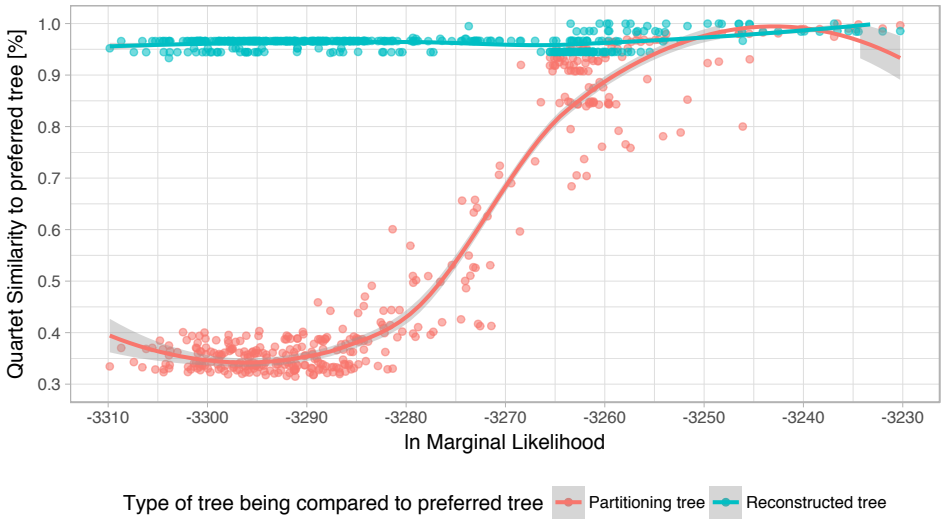
Any random tree shares a third of quartets with the true tree. While all quartets are equally informative in terms of topology, they clearly vary in how informative they are for partitioning. As trees get more and more dissimilar from the preferred tree, model fit (here a proxy for the suitability of the tree for partitioning) continues to decrease even though quartet similarity remains at 33%. Although new correct quartets are created by progressive changes to the tree, the overall order of tips and the sequences of branching events are disturbed. But it is exactly this order that is critical to the correct calculation of the number of steps on the tree and thus the calculation of accurate homoplasy values. Random trees in which the order of tips and the sequence of divergence events happen to be similar to the true tree thus lend themselves better to homoplasy partitioning.



(A) Effects of partitioning tree perturbation on topology in dataset HYO.



(B) Effects of partitioning tree perturbation on topology in dataset SCO.



(C) Effects of partitioning tree perturbation on topology in dataset THER.

FIGURE 4.1: Caption on next page.

(Previous page) Each homoplasy-partitioned model contributes two points to this figure displaying the relationship between the model's fit, the proximity of the partitioning tree to the preferred tree for a dataset, and the proximity of a reconstructed tree to the same preferred tree. Distance between trees is measured as quartet similarity (see section 3.1.2.5). Models based on a partitioning tree very similar to the preferred tree tend to show high marginal likelihoods, and model fit decreases as the partitioning tree becomes closer to random trees. In contrast, the reconstructed tree of a model remains quite similar to the preferred tree even when model fit is low.

4.1.3 Reconstructed and partitioning tree are independent

The accuracy of the reconstructed tree is not dependent on the accuracy of the partitioning tree. Bayesian analysis recovers a tree close to the preferred tree even when the partitioning tree is dissimilar (see Fig. 4.1). The quartet similarity of the reconstructed tree to the preferred tree is not correlated with the similarity of the partitioning tree to the preferred tree (Pearson's correlation coefficient $r = 0.232$).

The Bayesian reconstructed tree does not change much from analysis to analysis even as model fit decreases. Similarity to the preferred tree hovers around 95–100% for THER (Fig. 4.1c), 70–95% for SCO (Fig. 4.1b), and 90% for HYO (Fig. 4.1a). Even models partitioned on random partitioning trees infer reconstructed trees that are very similar to the preferred tree, sharing up to 95% of quartets.

4.1.3.1 Interpretation

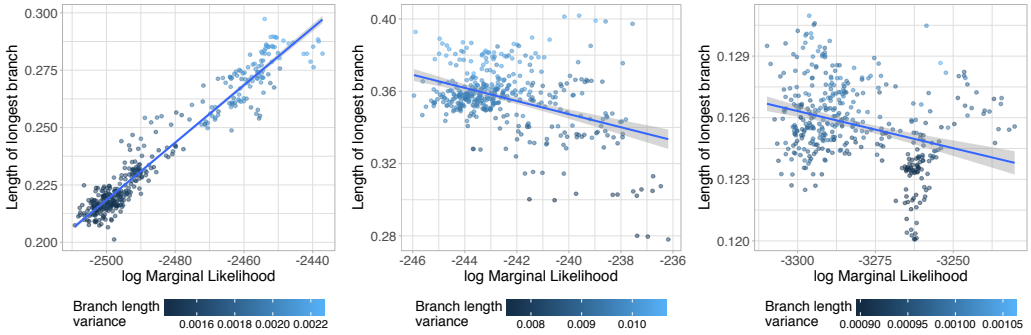
Uninformative partitioning does not appear to be sufficient to derail Bayesian inference of topology. After all, partitioning characters badly does not take away data from the model. The model is not told to expect certain evolutionary rates of each partition, merely to estimate their rates on separate distributions. If these distributions have the same shape, the partitions do not reflect different rate categories. In that case, the rate distributions will all look similar to the single distribution of an unpartitioned model, and between-partition rate multipliers will be close to 1. A badly partitioned model is thus able to infer the tree topology just as well as if there was no partitioning, but does so using more parameters and is penalised for this relative to more naïve models.

4.1.4 Estimation of branch lengths

If topology is inferred consistently and, as far as is verifiable, correctly, trees may still differ in their branch lengths. Branch lengths make up a large proportion of the parameters that have to be estimated in the course of a Bayesian analysis (Ronquist, Huelsenbeck, and Teslenko (2011), p.61). Problems with branch length inference, whether caused by a difficult dataset or a misspecified model, may therefore have an impact on whether a "good" tree is inferred (Ekman and Blaalid, 2011; Rannala, Zhu, and Yang, 2012).

The dataset HYO exhibits a strong trend towards poorly-fitting models inferring shorter longest branch lengths (see Fig. 4.2a and caption). But this seems inconsistent with studies that relate short branch lengths as a criterion for well-fitting models (Rannala, Zhu, and Yang, 2012; Ekman and Blaalid, 2011).

Minimum branch lengths also show a trend, but the relationship is reversed: better models infer shorter shortest branches, though the differences in length are miniscule (Linear regression for HYO: $F = 1443.0$, $df = 398$, $r^2 = 0.78$, $p < 0.001$, data not figured). This trend is uniform across datasets, though it is weak for THER ($r^2 = 0.13$). The variance in branch lengths is smallest when longest branches are short. This indicates that the best models infer branches that are quite similar to each other in length and (usually) short.



(A) HYO: $F = 5892.0$, $df = 398$, $r^2 = 0.94$, $p < 0.001$ (B) SCO: $F = 75.2$, $df = 398$, $r^2 = 0.16$, $p < 0.001$ (C) THER: $F = 44.8$, $df = 398$, $r^2 = 0.10$, $p < 0.001$

FIGURE 4.2: Scatterplot of the lengths of the longest branch of each result tree of a homoplasy-partitioned model (gamma-distributed branch length prior) versus model fit. The trendline is a line of best fit with 95% confidence interval in gray. Subfigure captions contain test statistics of linear regressions. Points are coloured by the sample variance of branch lengths. Shorter branches are associated with lower variance in all datasets. In HYO, the maximum branch length is strongly correlated with model fit: poorly-fitting models infer shorter branch lengths. In THER and SCO the trend is reversed and quite weak ($r^2 < 0.2$).

4.1.4.1 Interpretation

The effect of partitioning on branch length inference may be larger when branches are longer and span more evolutionarily disparate groups, leading to more changes along each branch. This would explain why branch lengths and model fit are correlated so strongly for HYO, but do not show a particularly strong relationship in SCO and THER, which are more limited in taxonomic scope.

It might make a difference whether branch lengths are estimated for each partition separately or from all character data in combination. Rosa, Melo, and Barbeitos (2019) reported best results when branch lengths remained linked as per default in MrBayes and my analyses followed their precedent in this regard.

4.2 Character Structure

When characters are partitioned by homoplasy, some consistently receive high homoplasy values and are sorted into the partition containing unreliable characters. Others show few or no superfluous steps even across large samples of random trees. It appears likely that there are inherent differences in character structure between these groups.

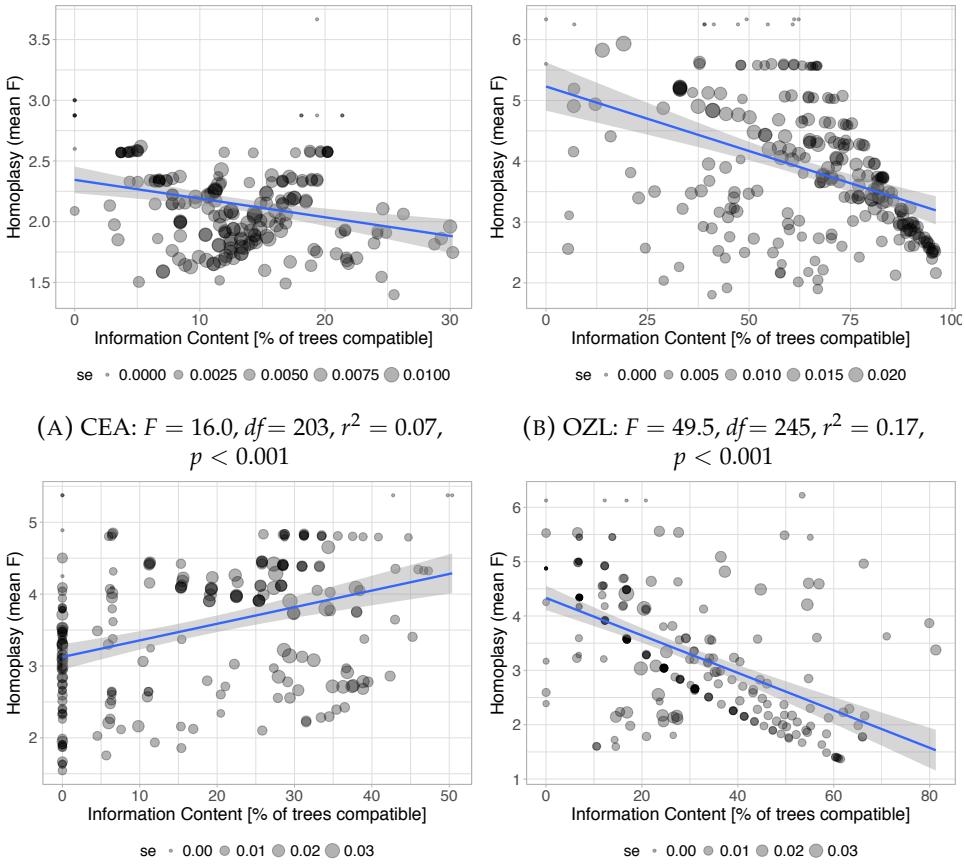
While a tree can only be built by combining information from several characters, it is possible to quantify the amount of phylogenetic information contained in a single character by studying its state distribution across taxa. I tested whether the tree-independent measure of Character Information Content (see section 3.4.5) is a suitable proxy for tree-dependent measures of homoplasy.

Fig. 4.3 shows mean values of Goloboff's f (see section 3.4.2) plotted against Information Content. High Information Content correlates weakly with low homoplasy for the majority of datasets (tested with linear regression, see captions in Fig. 4.3). While all tests were statistically significant, only in THER is this relationship statistically relevant ($r^2 = 0.30$). In HYO, the trend is reversed and the characters with the highest Information Content are most homoplastic.

4.2.1 Interpretation

Each dataset contains characters with an Information Content of 0%. These characters are compatible with all possible trees and therefore carry no information that helps to discern the true tree topology (parsimony uninformative characters). Often they range greatly in homoplasy, especially for HYO (Fig. 4.3c). Other characters with identical Information Content are also variably homoplastic, especially for THER and OZL (Fig. 4.3). Characters that are expected to be equally informative according to their IC thus show very different behaviours on trees.

It appears that Information Content and Goloboff's f do not capture the same information about a character. While there is some evidence that IC is picking up on an aspect of character structure, it is clear that there remains more to be discovered about how character state distributions relate to distributions of changes on a tree.



(A) CEA: $F = 16.0$, $df = 203$, $r^2 = 0.07$, $p < 0.001$

(B) OZL: $F = 49.5$, $df = 245$, $r^2 = 0.17$, $p < 0.001$

(C) HYO: $F = 35.4$, $df = 223$, $r^2 = 0.14$, $p < 0.001$

(D) THER: $F = 103.7$, $df = 240$, $r^2 = 0.30$, $p < 0.001$

FIGURE 4.3: Relationship between homoplasy and Character Information Content. All datasets but HYO exhibit lower homoplasy values for characters with high Information Content. Homoplasy was measured using Goloboff's f and averaged over 300 randomly sampled trees. Character Information Content was measured as the proportion of tree topologies compatible with a character out of all possible trees. The size of each marker corresponds to the standard error — smaller markers denote smaller errors. Trendline indicates line of best fit with shaded 95% confidence interval.

4.3 Synthesis

This chapter explored the effects of partitioning on tree topology and branch lengths.

I demonstrated that inference of topology is not hindered by uninformative partitioning, even though it depresses model fit (see section 4.1.3). If we are interested mainly in the relationships between different taxa and clades, it does not matter much whether the choice of partitioning tree is less than optimal — even a random tree can be sufficient.

The inference of branch lengths is more dependent on accurate estimates of evolutionary rates, and branch lengths did show dependence on how the model was partitioned (see section 4.1.4). The effect may be stronger for large datasets with cross-phylum taxon coverage like HYO (Fig. 4.2a). Thus if we are mainly interested in dating divergence times and key evolutionary events, it is more critical that the partitioning tree be reasonably close to the true tree.

We can never be entirely certain how accurately a tree represents the true evolutionary history of a clade. A tree-independent proxy for homoplasy could eliminate this source of uncertainty. The measure of Character Information Content aims to extract the relevant phylogenetic information from a character's state distribution. While Information Content and the homoplasy index Goloboff's f are weakly correlated, the link is not strong enough to make IC a viable tree-independent alternative to a homoplasy index.

Chapter 5

Results II — Partitioning strategies

This section presents the results of testing eight different partitioning strategies on five datasets. I briefly describe and visualise the data on the relative performance of the different strategies under the default branch length prior in MrBayes. Following on, I discuss what differences arise when partitioned analyses are run under other branch length priors.

5.1 Performance of different partitioning strategies

Models were ranked by their Bayes factor relative to the unpartitioned analysis within each dataset. I then took the arithmetic mean of the ranks of the eight partitioning strategies across analyses under the default branch length prior for all five datasets. The average ranks are:

1. Homoplasy partitioning based on preferred tree
2. Unpartitioned analysis
3. Homoplasy partitioning based on 300 random trees
4. Neomorphic-transformational partitioning with symmetric character state transition rates
5. tied:
   Partitioning by Information Content
   Neomorphic-transformational partitioning with asymmetric transition rates in the neomorphic partition
7. Randomly sorting characters into partitions
8. Partitioning by anatomy

The order of the partitioning strategies that received intermediate ranks is quite variable from dataset to dataset. The highest- and lowest-ranked strategies however are almost always the same. Homoplasy partitioning generally gives good results and includes the only partitioning strategy that consistently was preferable to not partitioning the data.

Fig. 5.1a summarises the model fit of different partitioning strategies under the default branch length prior for all five datasets. The range of marginal likelihoods varies strongly between datasets, with differences of up to 130 dB between the worst- and best-fitting models. This range is positively correlated with the number of taxa in each dataset, being largest for THER and OZL (130dB) and smallest for SCO ($<$ 10dB).

Figure 5.1b shows clearly that partitioning by anatomy and random partitioning give the worst results. They are outperformed by other strategies in almost all cases. Partitioning by homoplasy on the optimal tree is overall the most successful strategy, though it is in isolated cases outperformed by neomorphic-transformational partitioning or simply by an unpartitioned analysis (Fig. 5.1a). The unpartitioned model usually performs better than neomorphic-transformational partitioning, and partitioning by mean homoplasy over a sample of random trees is almost always preferable to partitioning by Information Content (Fig. 5.1b, see also section 4.2).

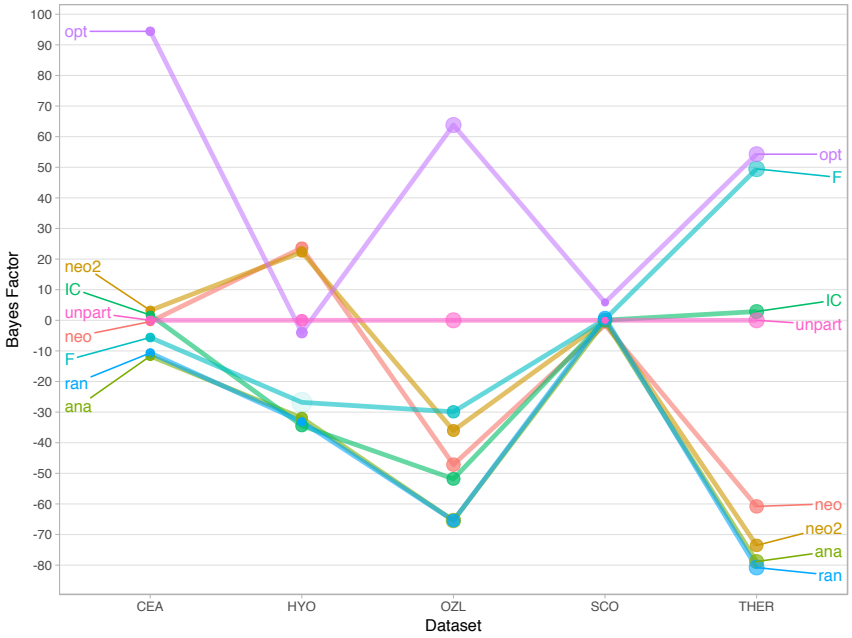
5.1.1 Neomorphic-transformational partitioning — symmetric or asymmetric transition rates?

Neomorphic-transformational partitioning operates on the premise that neomorphic and transformational characters evolve in fundamentally different ways (see section 3.4.3 and Sereno (2007) for definitions of character types). Neomorphic characters have a clear ancestral state and transitions from ancestral to derived states are more likely than reversals, leading to asymmetry in the initial state frequencies and the probabilities of transitioning between states.

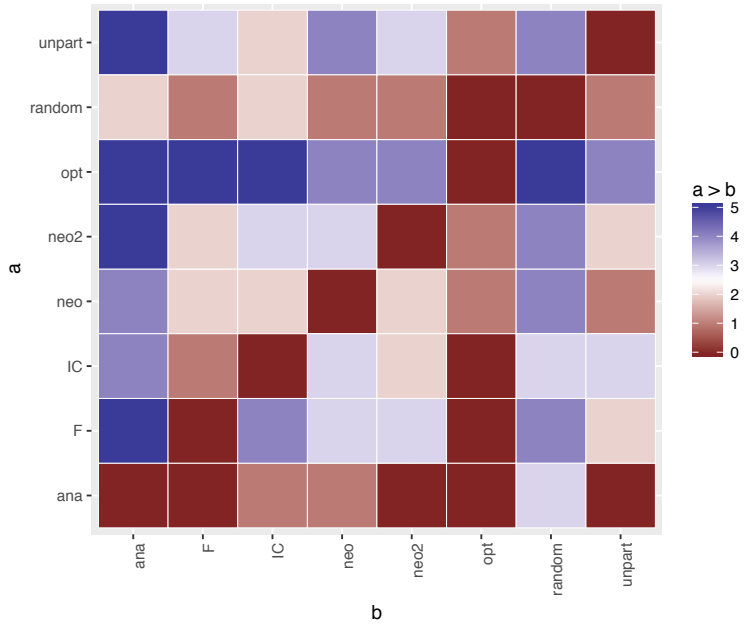
Neomorphic-transformational partitioning was tested twice with varied model settings. In **neotrans** analyses, I allowed state frequencies and transition rates to be asymmetric for characters in the neomorphic partition, while in **neotrans2** models they were symmetric for both transformational and neomorphic characters. Equal state frequencies performs slightly better in three cases, for datasets CEA, OZL, and SCO (see Table 5.2). Allowing for asymmetric transition rates in THER and HYO improved results by 13 and 1 dB respectively.

Neotrans models favour changes from ancestral to derived states and discourage reversals. State frequencies are estimated as a gamma dirichlet distribution with a single shape parameter that dictates the degree of asymmetry. If state frequencies and the linked transition rates are not fixed to be symmetric *a priori*, this shape parameter must be estimated, increasing the number of parameters that must be averaged over. If modifying the transition rates does not strongly improve inference, the model receives a lower marginal likelihood.

Neomorphic-transformational partitioning relies on correct assignment of the ancestral state in neomorphic characters. Errors in character formulation can easily prevent the success of this partitioning strategy.



(A) Bayes factors for all partitioning strategies across datasets.



(B) Heatmap showing pairwise comparisons of all eight partitioning strategies.

FIGURE 5.1: (A): Modified bumps chart of Bayes factors for all partitioning strategies across datasets. Point size and transparency are correlated with the size of the standard error — the larger the error, the more transparent and larger the point. (B): Heatmap showing pairwise comparisons of all eight partitioning strategies. This graph combines data from 5 datasets under the default branch length prior. Blue squares indicate that strategy **a** (row) tended to give better results than strategy **b** (column). Red quares indicate that strategy **b** was better on average. Partitioning strategy labels: **opt** = homoplasy partitioning (preferred tree); **neo2** = neomorphic-transformational part. (symmetric transition rates); **IC** = part. by Information Content; **unpart** = unpartitioned analysis; **neo** = neomorphic-transformational part. (asymmetric transition rates in neomorphic partition); **F** = homoplasy part. (300 random trees); **ran** = random part.; **ana** = part. by anatomy.

TABLE 5.1: Summary table of all partitioned models under the default gamma prior on branch lengths. Marginal likelihoods are presented in log units along with the model's Bayes factor relative to the unpartitioned model. Models are ranked within each dataset by their Bayes factor.

Dataset	Partitioning Strategy	ML [ln units]	BF [dB]	Rank
HYO	Unpartitioned	−2572.08	0	3
	Neomorphic-transformational	−2548.43	23.65	1
	Neomorphic-transformational 2	−2549.78	22.30	2
	Anatomy	−2603.99	−31.91	6
	Information Content	−2606.52	−34.45	8
	Homoplasy — random trees	−2598.88	−26.81	5
	Random partitions	−2605.21	−33.13	7
	Homoplasy — preferred tree	−2576.09	−4.01	4
CEA	Unpartitioned	−1596.83	0	4
	Neomorphic-transformational	−1597.25	−0.43	5
	Neomorphic-transformational 2	−1593.59	3.23	2
	Anatomy	−1608.51	−11.69	8
	Information Content	−1595.12	1.71	3
	Homoplasy — random trees	−1602.45	−5.62	6
	Random partitions	−1607.49	−10.67	7
	Homoplasy — preferred tree	−1502.40	94.42	1
OZL	Unpartitioned	−3742.16	0	2
	Neomorphic-transformational	−3789.31	−47.16	5
	Neomorphic-transformational 2	−3778.17	−36.01	4
	Anatomy	−3807.58	−65.42	7
	Information Content	−3793.97	−51.81	5
	Homoplasy — random trees	−3772.06	−29.90	3
	Random partitions	−3807.58	−65.42	7
	Homoplasy — preferred tree	−3678.36	63.80	1
SCO	Unpartitioned	−259.46	0	5
	Neomorphic-transformational	−260.96	−1.50	8
	Neomorphic-transformational 2	−259.59	−0.13	6
	Anatomy	−259.76	−0.29	7
	Information Content	−259.41	0.05	4
	Homoplasy — random trees	−259.09	0.37	3
	Random partitions	−258.71	0.76	2
	Homoplasy — preferred tree	−253.65	5.82	1
THER	Unpartitioned	−3700.92	0	4
	Neomorphic-transformational	−3761.72	−60.80	5
	Neomorphic-transformational 2	−3774.47	−73.55	6
	Anatomy	−3779.76	−78.83	7
	Information Content	−3698.06	2.87	3
	Homoplasy — random trees	−3651.45	49.48	2
	Random partitions	−3781.71	−80.79	8
	Homoplasy — preferred tree	−3646.65	54.28	1

TABLE 5.2: Effect of implementing symmetric or asymmetric state frequencies and transition rates under neomorphic-transformational partitioning. Negative Bayes factors indicate that model **neotrans2** (symmetric transition rates) performed better.

Dataset	Branch length prior	lnML [ln units]		BF [dB]
		neotrans	neotrans2	
HYO	default	−2548.43	−2549.78	1.35
	exponential	−2974.29	−2855.98	−118.31
	clock	−2581.41	−2580.62	−0.79
CEA	default	−1597.25	−1593.59	−3.66
	exponential	−1809.23	−1773.88	−35.35
	clock	−1617.92	−1614.75	−3.17
OZL	default	−3789.31	−3778.17	−11.14
	gamma	−3788.94	−3778.48	−10.46
	clock	−3837.81	−3821.56	−16.25
SCO	default	−260.96	−259.59	−1.37
	exponential	−278.01	−280.88	2.87
	gamma	−259.94	−259.27	−0.67
THER	default	−3761.72	−3774.47	12.75
	gamma	−3760.45	−3774.16	13.71
	clock	−3756.89	−3770.35	13.46

5.1.2 Number of partitions in anatomy partitioning

With every added partition, the number of parameters to be estimated increases. Partitioning by anatomy is especially prone to producing a large number of partitions. To test whether overparameterisation limited the performance of anatomy partitioning, I divided the datasets HYO and OZL first into eight and seven, then into four and six partitions.

Reducing the number of partitions improved model fit in all cases. The models with fewer partitions received higher marginal likelihoods, but typically only by a small margin (see Table 5.3). Halving the number of partitions for HYO led to a marked increase in model fit (BF=4 under the default branch length prior, higher under alternative priors). The effect for OZL was more modest (BF=1) since the number of partitions decreased less strongly. In several instances the relative order of partitioning strategies changed, with models partitioned by anatomy now performing better than randomly partitioned models, but this did not affect the overall mean ranks.

These results indicate that it is better to form fewer partitions when partitioning by anatomy. The added partitions do not aid inference sufficiently to warrant the estimation of so many added parameters.

TABLE 5.3: Effect of lowering the number of partitions in anatomy partitioning for datasets OZL and HYO. Positive Bayes factors indicate that the model with fewer partitions performed better.

(A) HYO

Prior	lnML [ln units]		BF [dB]
	8 Partitions	4 Partitions	
default	−2607.48	−2603.66	3.82
exponential	−2668.20	−2637.26	30.94
clock	−2636.61	−2631.50	5.11

(B) OZL

Prior	lnML [ln units]		BF [dB]
	7 Partitions	6 Partitions	
default	−3807.42	−3806.42	1.00
gamma	−3810.53	−3808.10	2.43
clock	−3853.87	−3849.73	4.14

5.1.3 Circularity in partitioning

The leading principle of Bayesian analysis is the separation of prior beliefs from evidence, so a prior should be in no way influenced by the data. Some of the partitioning strategies tested here violate this paradigm by estimating some property of the data before feeding that result into the model. To partition by homoplasy or by Information Content, we first need to calculate the homoplasy index or IC of each character using the character data. Homoplasy partitioning additionally requires a tree topology for this step. PartitionFinder2's strategy for choosing a partitioning scheme is also informed by the data since it calculates the maximum likelihood of each competing scheme given a tree topology. These methods all present modifications to the standard Bayesian paradigm, perhaps aligning them more closely with Empirical Bayes procedures where priors can be influenced by the data or even estimated directly from the data (Casella, 1985). Empirical Bayes approaches have been widely used (Carlin and Louis, 2000), for example in crop yield models (Green, Strawderman, and Thomas, 1992), to detect incorrect electricity meter readings (Rodrigues et al., 2019), and in the estimation of extinction times from the fossil record (see Hayes et al. (2020) and Alroy (2014), and discussion in Solow (2016) and Alroy (2016)).

The poor performance of partitioning by Information Content shows that simply incorporating some information from the data into the prior doesn't magically improve the model. Partitioning by homoplasy on the other hand produces models with consistently high fit to the data, creating the suspicion that model fit might be inflated. It appears possible that this circularity could be used to artificially improve the fit of such a model, whether inadvertently or purposefully.

To investigate whether a model partitioned by homoplasy on a preferred tree always fits the data better than the model which generated the preferred tree, I compared the marginal likelihoods of both models (see Table 5.4). The preferred trees for CEA and OZL were estimated with Bayesian models, with the CEA model using anatomy partitioning. The SCO, HYO, and THER preferred trees were constructed under maximum parsimony analysis, though the HYO paper also included node support values from a Bayesian analysis partitioned into neomorphic and transformational characters.

Only the CEA paper provided model fit values, so I modified the nexus files provided in the electronic supplementary material for HYO and OZL from the original MC^3 analyses to Steppingstone analyses and ran them in MrBayes to obtain marginal likelihoods. At first I increased the number of generations to the number of steps plus an additional burnin step times the number of generations in the original MC^3 analysis. However, these analyses would have exceeded Hamilton's maximum runtime of 72 hours, so for OZL I decreased the number of generations again by a factor of 10 from 1,640,000,000 to 164,000,000, while also increasing the sampling frequency by a factor of 10 from 4000 to 400 so that the same number of samples would be drawn. For HYO the number of generations of the modified analysis was 150,000,000. I also increased the burnin proportion from 10% to 30% and decreased the number of chains per run from 8 to 6, but kept the sampling frequency at 500. The OZL analysis showed good convergence, but unfortunately the HYO analysis repeatedly failed to converge and its results are not included in Table 5.4.

SCO, CEA, and OZL were also analysed by Rosa, Melo, and Barbeitos (2019) under partitioning by anatomy, homoplasy, and PartitionFinder2. The best model for each dataset used homoplasy partitioning on a maximum parsimony tree. The marginal likelihoods of these three models are included in Table 5.4.

From the limited data available, it appears that whether the homoplasy-partitioned model performs better than the original model is dependent on the dataset. The

TABLE 5.4: Bayes factor comparison of the fit of a model partitioned by homoplasy on a preferred tree (homoplasy model), versus the best Bayesian model from the paper that published that preferred tree (original model). Fit values are also compared to the best models from Rosa, Melo, and Barbeitos (2019), which were partitioned by homoplasy on a maximum parsimony tree.

Data	**Original model**	**Homoplasy model**		**Rosa et al. (2019)**	
	ML [ln units]	ML [ln units]	BF [dB]	ML [ln units]	BF [dB]
CEA	−1489.2	−1502.40	−13.2	−1562.01	−72.81
OZL	−3758.86	−3678.36	80.50	−3672.74	86.12
SCO	—	−253.65	—	−271.65	—

original anatomy-partitioned model for CEA shows better fit to the data than either homoplasy-partitioned model. It is worth mentioning that Rosa, Melo, and Barbeitos (2019) were unable to reproduce the success of CEA's anatomy partitioning, perhaps due to differences in parameter settings.

Another possible concern with homoplasy partitioning is the effect of the partitioning tree on the reconstructed tree. In Chapter 4 I tested homoplasy-partitioned model based on not just the preferred tree, which is presumed to be largely true, but also further removed perturbed trees and completely random trees. I demonstrated that homoplasy-partitioned models reconstruct tree topologies that closely resemble the preferred tree for a dataset, and not the partitioning tree on which homoplasy values were calculated. Any hidden influence that the preferred tree could have upon the reconstructed topologies gets "lost in translation" when characters are sorted into their homoplasy categories. Thus I consider partitioning by homoplasy to be no more circular than other data-conscious approaches to partitioning.

5.1.4 Interpretation

Partitioning by homoplasy emerged as the best partitioning strategy overall. It requires a partitioning tree reasonably close to the true tree, but as demonstrated in section 4.1 this tree does not need to be highly accurate. Conceived as a tree-independent alternative to homoplasy partitioning on a preferred tree, partitioning by Information Content was unable to extract the necessary information about evolutionary rates from the raw character data. The method of calculating mean homoplasy over a large sample of random trees was similarly unsuccessful.

Neomorphic-transformational partitioning, or partitioning by character type, gave only mediocre performance. Modelling the asymmetry in transition rates and state frequencies expected in neomorphic characters can improve model fit, but this is highly dependent on the correct assignment of ancestral states (see section 5.1.1). Since ancestral state reconstruction is always dependent on a phylogenetic tree, this polarity poses a problem when using ancestral states for phylogenetic inference. Pagel and Lutzoni (2002) suggest reconstructing ancestral states on each tree from an MCMC sample of trees and accepting the consensus as the evolution of that character. This method is indeed tree-independent, however it still relies on the correct specification of the Bayesian model used to collect the MCMC sample. The challenges to accurate ancestral state reconstruction are thus very much the same as to reconstruction of phylogenies in general, and correct ancestral states should not be a critical requirement for phylogenetic inference.

Random partitioning of characters was included to provide a benchmark of minimum performance; the other partitioning strategies were definitely expected to beat this bar. Surprisingly, this was not the case for partitioning by anatomy which received worse model fit values than random partitioning in several cases.

This provides positive evidence against morpho-functional correlation among characters, at least at the scales of the datasets studied here. However, previous studies did find evidence of mosaic evolution. Felice and Goswami (2018) found that regions of the avian cranium discriminated by different developmental origins evolved at varying rates. Being exclusively based on morphometric scans of bird skulls, their dataset was more limited in the range of characters than the avian datasets used in this study, OZL and CEA. Linkage between characters caused by mosaicism may readily occur at such small spatial scales, especially if characters are not independent as is likely with morphometric data.

The original CEA model as well as the best anatomy-partitioned models of Rosa, Melo, and Barbeitos (2019) had unlinked branch lengths, whereas I kept branch lengths linked across partitions. If anatomical partitioning always works better when branch lengths are estimated independently for each partition, this would explain why my anatomy-partitioned models always performed more poorly than their unpartitioned counterparts. It is also possible that partitioning by function rather than anatomical region (as in the original study of the dataset CEA (Clarke and Middleton, 2008)) would have been more successful.

In general, it appears favourable to limit the number of partitions (see section 5.1.2). Thus, partitioning off only one or two functional groups of interest and grouping the remaining characters together also might have yielded better results.

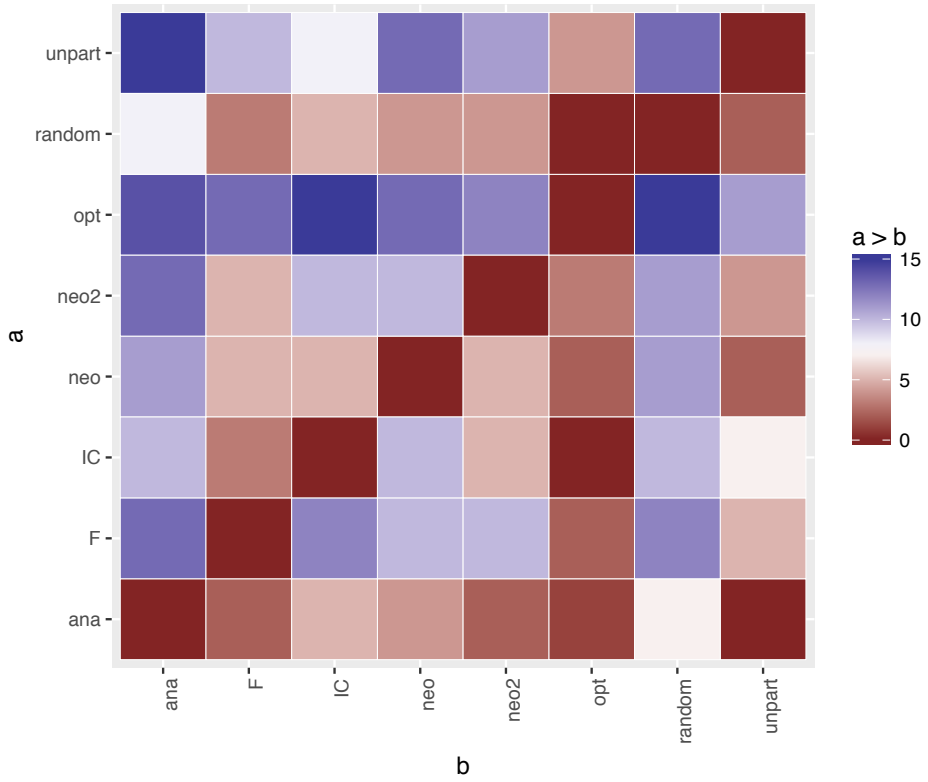
5.2 Relative performance of branch length priors

In addition to the different partitioning strategies, I also tested various methods of inferring branch lengths. Apart from the default gamma dirichlet prior with shape parameter $\alpha_T = 1$, I tested a gamma dirichlet prior with fitted α_T, a fitted exponential prior and a relaxed clock model (see section 3.2.1.3).

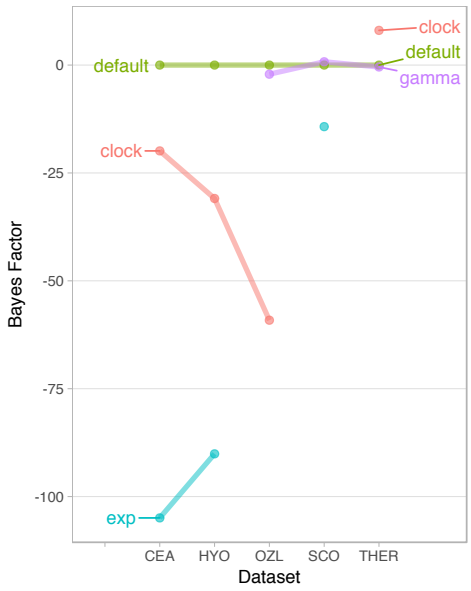
Figure 5.2a compares the relative performance of all partitioning strategies in pairs. The overall result remains the same when incorporating data from all branch length priors, not just the default prior (cf. Fig. 5.1b). Homoplasy partitioning based on the optimal tree still performs best, followed by unpartitioned analyses. Partitioning by anatomy and random partitioning give the worst results.

Figure 5.2 compares results for each datasets under the different branch length priors. The Bayes factor for a dataset-prior combination varies by up to 300 dB between partitioning strategies (compare e.g. HYO under the exponential prior in Figures 5.2b and 5.2c). Nevertheless, the overall pattern is remarkably consistent over different partitioning strategies.

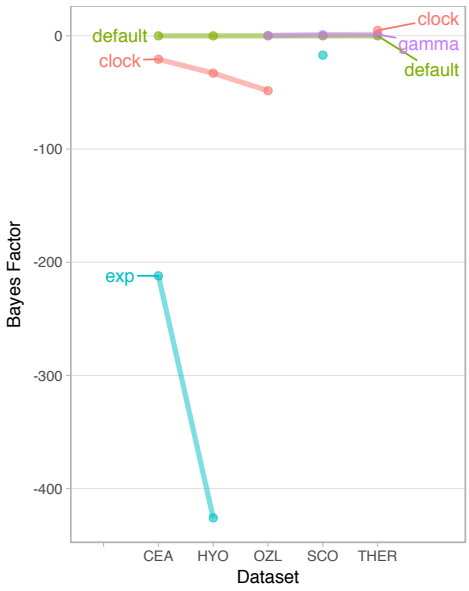
The exponential prior produces by far the poorest models for every dataset it was tested on, and under every partitioning strategy. For dataset HYO partitioned into neomorphic and transformational characters, using an exponential branch length prior is over 400 dB worse than using the default prior. It was the only prior to introduce substantial variation in the relative ranks of different



(A) Heatmap showing pairwise comparisons of all eight partitioning strategies.



(B) Fit of unpartitioned models



(C) Fit of neomorphic-transformational models

FIGURE 5.2: Comparison of model fit under different branch length priors. (A) Heatmap showing pairwise comparisons of all eight partitioning strategies. Includes 15 data points for each comparison as this graph combines data from 5 datasets and 3 branch length priors. (B, C) Graphs of unpartitioned model and model with neomorphic-transformational partitioning. Each plot shows Bayes factors for a single partitioning strategy, separated along the x-axis by dataset and colour-coded according to the branch length prior. Bayes factors were calculated relative to the default prior. Branch length prior key: **default** = default gamma dirichlet prior; **exp** = fitted exponential prior; **gamma** = fitted gamma dirichlet prior; **clock** = relaxed clock model. Prior specifications are found in section 3.2.1.3.

partitioning strategies. For HYO, it made the anatomy partitioned model perform much better than usual, but not better than the unpartitioned model which was the highest-ranked model overall. Neomorphic-transformational partitioning ranks highest under the default prior, and lowest under the exponential prior.

For CEA, neomorphic-transformational partitioning performed much more poorly, while the models partitioned by information content and homoplasy on random trees were ranked higher than under the other priors.

For SCO, the unpartitioned model was ranked first, and neomorphic-transformational partitioning performed better than under alternative priors.

The clock model generally performed slightly worse than the default prior. The only exception is the dataset THER, where the clock model yields the best results, sometimes by a margin of several dB (see Fig. 5.2b).

The fitted gamma dirichlet prior performs very similarly to the default static gamma dirichlet prior. However, the Bayes factors between these models are too small to carry much significance under Bayesian model testing.

5.2.1 Interpretation

According to the MrBayes manual (Ronquist, Huelsenbeck, and Teslenko, 2011), clock models can be expected to win out over non-clock models in a Bayes factor test. Clock models are a special case of non-clock models and are therefore expected to infer just as good estimates of phylogeny. Since Bayesian model comparison takes into account the number of estimated parameters when assessing model fit, a model with more parameters that gives just as good results as another will receive lower model fit, because the added complexity of the model does not translate to an increase in inferential power (see section 2.3.2.1). In a strict clock model for n taxa, n independent node times are estimated, whereas in a non-clock model the number of estimated branch lengths is closer to $2n$. A relaxed clock model employs an intermediate number of parameters: Specifically, the relaxed clock with independent gamma rates that was used in this study estimates n node times plus one additional parameter per partition. Thus, if the clock model succeeds in estimating the phylogeny just as well as the more parameter-rich non-clock model, the clock model has a higher marginal likelihood. Considering this natural advantage of clock models, it is surprising that my models using a relaxed clock did not give better results than those with the default prior.

Exponential models show by far the worst model fit. As unconstrained branch length models, they are expected to be worse than clock models due to the additional parameters, but they also perform much more poorly than the default and gamma models. The exponential distribution with $\lambda = 1$ was the default branch length prior until MrBayes v3.2.2 (Ronquist, Huelsenbeck, and Teslenko, 2011). After Rannala, Zhu, and Yang (2012) and Zhang, Rannala, and Yang (2012) reported that a gamma dirichlet prior gave much better results, the default was changed to a gamma dirichlet distribution with shape parameter $\alpha_T = 1$. Gamma dirichlet

priors are less prone to inferring overly long branches, thus avoiding long branch attraction and clustering of extant tips on trees combining fossil and living taxa (Ronquist, Lartillot, and Phillips, 2016).

Fitting the prior to each dataset posed no significant improvement to model fit compared to using the default prior. It may have an impact on branch lengths, however this was not tested here. Strictly based on model fit there is no evidence that fitting the prior leads to better inference of phylogeny.

5.3 Synthesis

The aim of this chapter was to establish which partitioning strategy and which branch length prior are most suitable for morphological characters. I presented evidence relating to their performance in Bayesian inference and discussed challenges to each strategy's implementation.

In most cases, the relative performance of a partitioning strategy does not change depending on the branch length prior — only the absolute fit of the combined model is affected. Partitioning by homoplasy on a reasonably accurate tree is the best partitioning strategy overall, while partitioning by anatomy performs poorly and cannot be recommended. Neomorphic-transformational partitioning has the potential to perform well given the right conditions, but requires knowledge of ancestral states and is thus unsuitable unless character polarity can be clearly established. No alternative branch length prior tested here presented an improvement over the default setting. Fitting the prior to the dataset does no harm, but may be unnecessary and can be safely omitted.

Chapter 6

Discussion

6.1 What is the best partitioning strategy . . .

The idea behind partitioning is that it allows us to provide the model with information about the variation in evolutionary rates in a dataset. Phylogenetic datasets inevitably contain some rate variation, and in some cases it is straightforward to predict — for example, coding DNA sequences are under stricter mutation control and evolve more slowly than non-coding sequences. It is much more difficult to predict the evolutionary rate of a morphological character.

6.1.1 . . . for model fit?

After testing a range of new and previously proposed partitioning criteria on several published datasets, partitioning by homoplasy on a reasonably well-fitting tree emerges as the best strategy.

Weighting by homoplasy has been employed in Maximum Parsimony analyses for decades and is — while not entirely uncontroversial, see e.g. Kluge (1997) — generally accepted to work well, if the concavity constant k suits the dataset. Weighting and partitioning by homoplasy operate on the premise that we should divide characters according to their evolutionary rate. This seems to work very well using the measures of homoplasy at our disposal, such as Goloboff's f.

Rosa, Melo, and Barbeitos (2019) tested homoplasy-based partitioning against anatomical partitioning and partitioning using PartitionFinder2 (Lanfear et al., 2016). PartitionFinder2 mainly includes an algorithm designed to compare all possible ways of partitioning a dataset by forming partitions from user-defined groups of characters. It uses a single tree on which the likelihoods of all possible partitions are calculated. For each partitioning scheme, the likelihoods of the constituting partitions are summed and the best scheme is chosen via AIC, AICc, or BIC.

Partitioning by homoplasy produced better models than PartitionFinder2, suggesting that the algorithm was unable to identify the best-fitting partitioning strategy. The best partitioning scheme is chosen by information criterion, not by the Bayes factor. Perhaps this caused models with higher maximum likelihoods, rather than highest marginal likelihoods, to be selected. Rosa, Melo, and Barbeitos (2019) state that the scheme selected by PartitionFinder2 did not depend on which tree

was used, further confirming the disconnect between partitioning and tree topology.

I observed that lowering the number of partitions improved model fit under anatomy partitioning, but an increased number of partitions may be tenable if enough partitions are set up to eliminate the need for estimating within-partition ACRV (Among-Character Rate Variation). I limited the number of partitions under homoplasy-based partitioning to 3, while Rosa, Melo, and Barbeitos (2019) allowed as many partitions as the number of discrete homoplasy values. As many as 14 partitions were implemented for OZL. Rosa, Melo, and Barbeitos (2019) found that their highly partitioned homoplasy models performed best when branch lengths were linked and rates were assumed to be equal for all characters within a partition. They concluded that partitioning by homoplasy accounted for all the variation in evolutionary rates, eliminating the need for modelling within-partition variation.

Rosa, Melo, and Barbeitos (2019) used an exponential prior to estimate branch lengths, which has not been the recommended setting since 2013 (Ronquist, Huelsenbeck, and Teslenko, 2011; Zhang, Rannala, and Yang, 2012; Rannala, Zhu, and Yang, 2012). Despite the suboptimal prior, their best model for OZL showed higher model fit than mine, by a significant margin of 6 dB. But for CEA and OZL, my models with the default gamma dirichlet prior on branch lengths, 3 partitions, and gamma-distributed rate variation within partitions beat their by 60 dB and 18 dB.

Rosa, Melo, and Barbeitos (2019)'s results show that homoplasy is indeed a very accurate proxy for evolutionary rate, but in terms of model fit it is preferable to limit the number of partitions and model rate variation within partitions as well as between partitions.

6.1.2 . . . for inferring topology?

In my experiments, tree topology was largely unaffected by changes to the partitioning tree which drastically lowered model fit (see section 4.1.3). This is corroborated by Kainer and Lanfear (2015), who point out that large differences in model fit are not necessarily accompanied by substantial changes to tree topology. Kainer and Lanfear (2015) caution against indiscriminate application of partitioning, stating that while the partitioning scheme rarely has an influence on topology, for some datasets it did change the tree topology. However, the mean difference between topologies stated by them do not exceed the values that one might get by moving a single tip by a short distance (cf. section 1.4 of Smith (2019e)).

While it is known that reconstructed topologies differ between data partitions (Duchêne et al., 2011), if the topology is reconstructed from all data in combination there does not appear to be much variation. In reverse, Posada and Crandall (2001) also showed that model selection algorithms prefer the same models whether they are conditioned on several trees or a single topology. I therefore argue that tree topology is largely independent of partitioning. Broad taxon coverage and the inclusion of fossils, preventing long branch attraction (Graybeal, 1998), and a large

set of independent characters which have been formulated and coded following recommended best practice, preventing deep root attraction (Ronquist, Lartillot, and Phillips, 2016), are likely much more influential in ensuring that topology is inferred correctly.

6.1.3 . . . for inferring branch lengths?

Branch lengths are closely tied to evolutionary rates, since the length of a branch is typically measured in the number of changes expected along it. I found minor but significant variation in branch lengths depending on the fit of the model, i.e. the suitability of the partitioning method.

Marshall, Simon, and Buckley (2006) emphasise the importance of modelling between-partition rate variation through either unlinking branch lengths or a rate multiplier, because failing to account for this variation can cause overly long branches.

Upon varying the branch length prior, I found that this rarely has an effect on the relative performance of different partitioning strategies. The only prior where ranks regularly varied was the exponential prior. This prior has been shown to infer suboptimal branch lengths (Rannala, Zhu, and Yang, 2012; Zhang, Rannala, and Yang, 2012) and has been replaced by the more versatile gamma dirichlet prior. It seems likely that the limitations of the ill-fitting prior interfered with proper inference. The default gamma prior, the fitted gamma prior and the relaxed clock prior all largely agree that homoplasy partitioning performs best. The gamma priors achieve the highest model fits.

6.2 What can partitioning tell us about evolution?

With a little bit of biological knowledge, is it possible to predict the evolutionary behaviour of a character from its character formulation? If we can approximate evolutionary rates by homoplasy, it should also be possible to use other criteria for this. For example, it seems intuitive that transformational characters like plumage colouration, tooth length, or beak shape can change more quickly than two bones can fuse or a rodent can lose its tail in a more complex neomorphic evolutionary novelty. Similarly, when an ecological driver acts on an organism, it probably will effect change in a suite of characters related by function, not one single character. Characters in each suite may then evolve at the same rate since they are under the same selective influence.

Two of the partitioning strategies tested here incorporated these macroevolutionary hypothesis into their design. Partitioning by character type allowed neomorphic characters to evolve more slowly than transformational characters, and anatomy partitioning tested whether characters evolve at similar rates within morphofunctional groups. If these models with their additional parameters more closely resembled true evolutionary processes, we would expect to see the fit of the model improve in comparison to a more restrictive model of evolutionary rates.

6.2.1 Mosaic evolution

If anatomical partitioning were often successful, it would indicate that the evolutionary rates of characters are linked within morphofunctional groups. Such mosaic evolution has been oft-proposed and observed in several empirical datasets, but its utility in partitioning could not be confirmed here.

Mosaic evolution plays a role at small anatomical scales, such as in morphometric datasets like that of the avian skull in Felice and Goswami (2018), and could be used as a partitioning criterion here. To partition datasets describing the whole body, anatomy seems less suitable. Krause et al. (2014) also describe evidence of mosaicism in a Gondwanatherian skull which has both ancestral and highly derived features, but these do not appear to be grouped to distinct regions of the skull. Anatomical partitioning would thus be unable to assign these characters correctly.

Diogo, Molnar, and Wood (2017) report mosaicism in the internal anatomy of chimpanzees, bonobos, and humans, but the differences between the taxa are restricted to seven characters over three distinct regions. While the differences are clearly segregated by anatomical region, it is unlikely that the presence of these few characters would affect the average rate of evolution of anatomical partitions. Gaubert et al. (2005) found that anatomical partitions of data lend support to different clades on a parsimony tree of feliformian viverrids. This suggests punctuated rapid evolution in certain groups of characters, but not necessarily that rates differ consistently through time between the groups.

In contrast to my analyses, for Rosa, Melo, and Barbeitos (2019) anatomical partitioning with unlinked branch lengths was preferable to not partitioning the data. Their results indicate that anatomical partitioning works better when branch lengths are unlinked and thus estimated independently for each partition. If branch lengths are very disparate between anatomical partitions, keeping branch lengths linked would not allow the model to sufficiently account for this. Furthermore, Wright (2015) found that the anatomical partitions of Clarke and Middleton (2008) with linked branch lengths performed vastly better than any partitioning scheme suggested by PartitionFinder2 by. Though my comparisons did not include partitioning with PartitionFinder2, this stands in stark contrast with my observation that anatomical partitioning was equivalent to sorting characters into partitions randomly.

Indeed, Goloboff et al. (2018) argued that there is no logical connection between mosaicism and variation in evolutionary rate at all — mosaics of plesiomorphic and derived characters can just as well arise under uniform rates. A group of characters may evolve more quickly than another at one point in time, but at another time or in another lineage the rate in the other group may be higher. Averaged across the whole evolutionary time range covered by the phylogenetic tree the rates between different groups would be similar.

While mosaic evolution is a widespread pattern in diversification processes, its

connection to evolutionary rates and its utility in partitioning cannot be corroborated here.

6.2.2 Character type

While contingent coding (Sereno, 2007) is employed more and more in dataset recoding or assembly (Bielecki et al., 2014; Thompson et al., 2012), there is very little precedent for partitioning into neomorphic and transformational characters. To my knowledge, only Sun et al. (2018b) and Moysiuk and Caron (2019) have thus far employed neomorphic-transformational partitioning.

In my analyses, partitioning by character type generally presented no improvement over not partitioning the data. Furthermore, equal transition rates were preferable to allowing more frequent transition absent ⇒ present than reversals. This is corroborated by Gamble et al. (2012), who showed that counterintuitively, adhesive toepads of geckos were gained as frequently as they were lost. Swenson, Richardson, and Bartish (2008) also demonstrated repeated gains and losses of neomorphic characters on a plant phylogeny.

Even with equal transition rates, partitioning by character type is only successful in isolated cases. The HYO dataset assembled by Sun et al. (2018b) obeyed by the rules of contingent coding, which may explain the strategy's success here. The CEA dataset was not assembled with Sereno's instructions for character formulation in mind and contains several characters unsuited to this partitioning strategy, but nevertheless model fit improves when it is partitioned into these two character types. However, these isolated instances of success are not sufficient to recommend neomorphic-transformational partitioning for every dataset.

6.2.3 Verdict

In general, homoplasy appears to be the only reliable estimator of evolutionary rate for morphological data available to us currently. Morphological evolution remains unpredictable and spurious. While mosaic evolution demonstrably occurs, it does not aid in phylogenetic inference. The distinction by character type into neomorphic and transformational characters similarly does not seem to have implications for the evolutionary rates of different characters.

One shortcoming of homoplasy as a partitioning criterion, the dependency on a partitioning tree, may be circumvented in the future with the invention of a tree-independent proxy for homoplasy. Such a proxy might be calculated directly from the character data, possibly by looking for correlations of character states between taxa, character state proportions, or some more derived method. Information Content (IC) attempts to capture information about homoplasy in a character by calculating the proportion of topologies on which it is compatible (see section 3.4.5), but models partitioned by IC perform much worse than those partitioned by homoplasy. However, the topology of the partitioning tree does not appear to have a

substantial influence on the tree reconstructed by the partitioned model, and model fit is not impacted much as long as the tree is non-random.

Partitioning by homoplasy fulfills my criteria for a good partitioning strategy:

Equal partitions — It allows control over the number of partitions and the size of each partition, preventing partitions from becoming so small that parameters cannot be estimated based on too little data.

No specialist knowledge — It can be applied to a dataset without having specialist anatomical knowledge of a clade.

Partitions are reproducible — Characters are not assigned to partitions based on a criterion that must be interpreted by the taxonomist, but simply by a value.

Partitioning by homoplasy on a reasonable tree is therefore the only partitioning strategy that I can recommend to other workers in phylogenetics.

Chapter 7

Conclusions and suggestions for further work

7.1 Conclusions

I demonstrated that homoplasy-based partitioning is a good proxy for evolutionary rates and is not misled by mismatches between the partitioning tree and the true tree. Partitioning does not have a major effect on topological inference, but does influence branch lengths. An optimal topology can thus be inferred without partitioning. Homoplasy partitioning on this reconstructed tree topology should then be able to infer accurate branch lengths. Under linked branch lengths, the branch length prior implemented as the default in MrBayes v3.2.6 was found to perform as well as or better than the alternative priors tested.

I further showed that neither anatomical location nor character type allow us to predict the evolutionary rate of a character. Evolutionary rates of morphological characters are linked neither within suites of characters evolving in tandem, nor to the complexity or novelty of the character. While mosaic evolution demonstrably occurs, it cannot predict rates of change over evolutionary timescales.

7.2 Further work

7.2.1 Neomorphic-transformational partitioning

The partitioning strategies I tested included two variations of neomorphic-transformational partitioning. Any properly formulated character can be classified as neomorphic or transformational, depending on whether it describes the evolution of a new feature or a change in an existing trait. However, the majority of datasets used here, and indeed most published character lists, include combination characters. They state the absence or presence along with transformational states of a character. I assigned such characters to the transformational partition.

Future analyses could instead recode combination characters into a neomorphic and a transformational character. This would constitute an improvement to the accuracy of partitioning through the elimination of neomorphic aspects in

characters in the transformational partition, and may improve the performance of neomorphic-transformational partitioning.

7.2.2 Ontological partitioning

Partitioning characters based on developmental stage is particularly applicable to organisms with clearly demarcated life stages, such as many arthropods. The justification of ontology partitioning lies in the fact that different life stages of the same animal are subject to completely different ecological pressures. Since evolution occurs under the constraints of ecological drivers, characters of the larval stage could evolve at different rates to those of the adult animal. Partitioning by developmental stage was not applicable to all datasets used in this study and thus was omitted, but could be tested in the future.

7.2.3 The number of partitions

The number of partitions varied between partitioning strategies, but was arbitrarily limited to three for most, including homoplasy partitioning. I found that decreasing the number of partitions improved model fit for partitioning by anatomy, indicating that lower numbers of partitions are favourable under ill-fitting criteria. I did not explore the effect of varying the number of partitions under a strategy that reliably predicts evolutionary rate, such as homoplasy partitioning.

7.2.4 A future model of morphological evolution

The development of a novel model of morphological evolution is inevitable in the near future. This model will be able to accommodate directional evolution, account for interdependence between morphological characters due to a shared genetic basis, and recognise convergence due to common ecological pressures (Ronquist, Lartillot, and Phillips, 2016). It will model correlation between morphological and molecular clocks and accommodate rate variation among characters.

I hope that this thesis can play a small part in furthering our understanding of morphological evolution and aiding the development of this new model.

Appendix A

Partitions and Character Ordering

A.1 Partitions

A.1.1 HYO

Neomorphic-transformational partitions:

neomorphic - 129
1 6 7 9 10 15 18 19 20 21 22 24 26 29 30 34 35 37 38 40
43 44 45 48 49 53 54 55 56 57 58 59 60 61 65 66 68 69
71 76 78 79 80 81 83 86 87 88 91 92 93 94 97 98 99 100
101 102 106 108 109 111 118 119 121 122 125 126 129 136
137 138 139 140 142 143 144 145 148 150 151 153 154 155
156 157 158 160 164 168 170 174 175 176 177 179 180 181
182 183 184 185 186 187 188 190 192 193 196 199 202 203
204 205 206 207 209 210 211 212 214 215 216 217 218 219
220 222 225;

transformational - 96
2 3 4 5 8 11 12 13 14 16 17 23 25 27 28 31 32 33 36 39
41 42 46 47 50 51 52 62 63 64 67 70 72 73 74 75 77 82
84 85 89 90 95 96 103 104 105 107 110 112 113 114 115
116 117 120 123 124 127 128 130 131 132 133 134 135 141
146 147 149 152 159 161 162 163 165 166 167 169 171 172
173 178 189 191 194 195 197 198 200 201 208 213 221 223
224;

4 anatomical partitions:

Soft exterior - 42
207 208 209 210 211 212 35 36 37 38 39 20 21 22
23 24 25 162 163 164 165 166 167 168 172 173 176
177 213 214 26 27 28 29 30 31 32 33 34 174 175 215;

Feeding and Digestion - 55
40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55
56 57 58 59 60 61 62 63 64 178 179 180 181 182 183
184 185 186 187 188 189 190 191 192 193 194 195
196 197 198 199 200 201 202 203 204 205 206 216;

Reproductive & Sensory - 44 141 142 143 144 145 146 147 148 149 150 151 152
153 154 155 156 157 158 159 160 161 9 10 11 12 13
14 15 16 17 18 19 169 170 171 217 218 219 220 221
222 223 224 225;

Shell - 84 1 2 3 4 5 6 7 8 65 66 67 68 69 70 71 72 73 74 75
76 77 78 79 80 81 82 83 84 85 86 87 88 89 90 91
92 93 94 95 96 97 98 99 100 101 102 103 104 105
106 107 108 109 110 111 112 113 114 115 116 117
118 119 120 121 122 123 124 125 126 127 128 129
130 131 132 133 134 135 136 137 138 139 140;

A.1.2 CEA

Neomorphic-transformational partitions:

neomorphic - 68 2 3 4 5 6 8 11 14 18 19 20 25 26 29 34 39 40 41 42 46
50 51 52 54 60 61 65 68 71 74 75 78 80 83 88 92 94 95
98 109 111 112 117 120 125 126 129 130 131 132 133 136
138 147 148 154 156 165 166 167 172 175 177 181 184 192
194 204;

transformational - 137 1 7 9 10 12 13 15 16 17 21 22 23 24 27 28 30 31 32 33
35 36 37 38 43 44 45 47 48 49 53 55 56 57 58 59 62 63
64 66 67 69 70 72 73 76 77 79 81 82 84 85 86 87 89 90
91 93 96 97 99 100 101 102 103 104 105 106 107 108 110
113 114 115 116 118 119 121 122 123 124 127 128 134 135
137 139 140 141 142 143 144 145 146 149 150 151 152 153
155 157 158 159 160 161 162 163 164 168 169 170 171 173
174 176 178 179 180 182 183 185 186 187 188 189 190 191
193 195 196 197 198 199 200 201 202 203 205;

5 anatomical partitions:

Skull - 52 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22
23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41
42 43 44 45 46 47 48 49 50 51 52;

Anterior trunk - 45 53 54 55 56 57 58 59 60 61 70 71 72 73 74 75 76 77 78 79
80 81 82 83 84 85 86 87 88 89 90 91 92 93 94 95 96 97 98
99 100 101 102 103 104 105;

Posterior trunk - 25 62 63 64 65 66 67 68 69 155 156 157 158 159 160 161 162
163 164 165 166 167 168 169 170 171;

Forelimbs - 49 106 107 108 109 110 111 112 113 114 115 116 117 118 119
120 121 122 123 124 125 126 127 128 129 130 131 132 133

134 135 136 137 138 139 140 141 142 143 144 145 146 147
148 149 150 151 152 153 154;

Hindlimbs - 34 172 173 174 175 176 177 178 179 180 181 182 183 184 185
186 187 188 189 190 191 192 193 194 195 196 197 198 199
200 201 202 203 204 205;

Ordered characters:

ctype ordered 1 9 12 24 32 53 55 62 63 67 69-70 72 77 79 82 106 114 119 141
144 151 155 160 162 173 178 180 183 185 188 191 195-199 205;

A.1.3 OZL

Neomorphic-transformational partitions:

neomorphic - 100 7 13 15 16 17 18 20 21 22 26 33 34 35 37 40 41 42 45
46 47 48 50 52 54 55 59 63 65 68 71 76 77 78 81 88 89
90 91 93 98 101 102 113 114 116 119 120 123 124 125 126
127 130 133 135 136 137 138 147 149 150 151 157 158 164
167 170 172 173 174 175 179 180 182 184 186 187 189 190
191 192 195 196 198 199 200 206 207 208 210 220 224 230
231 238 240 242 244 245 246;

transformational - 147 1 2 3 4 5 6 8 9 10 11 12 14 19 23 24 25 27 28 29 30 31
32 36 38 39 43 44 49 51 53 56 57 58 60 61 62 64 66 67
69 70 72 73 74 75 79 80 82 83 84 85 86 87 92 94 95 96
97 99 100 103 104 105 106 107 108 109 110 111 112 115
117 118 121 122 128 129 131 132 134 139 140 141 142 143
144 145 146 148 152 153 154 155 156 159 160 161 162 163
165 166 168 169 171 176 177 178 181 183 185 188 193 194
197 201 202 203 204 205 209 211 212 213 214 215 216 217
218 219 221 222 223 225 226 227 228 229 232 233 234 235
236 237 239 241 243 247;

6 anatomical partitions:

Skull & Mandible - 50 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21
22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39
40 41 42 43 44 45 46 47 48 246 247;

Vertebrae & Ribs - 33 49 50 51 52 53 54 55 56 57 58 59 60 61 62 63 64 65 66
67 68 69 70 71 72 73 74 75 76 77 78 79 80 81;

Thorax & Sternum - 39 82 83 84 85 86 87 88 89 90 91 92 93 94 95 96 97 98 99
100 101 102 103 104 105 106 107 108 109 110 111 112
113 114 115 116 117 118 119 120;

Forelimbs - 57 121 122 123 124 125 126 127 128 129 130 131 132 133
134 135 136 137 138 139 140 141 142 143 144 145 146
147 148 149 150 151 152 153 154 155 156 157 158 159
160 161 162 163 164 165 166 167 168 169 170 171 172
173 174 175 176 177;

Pelvic girdle - 23 178 179 180 181 182 183 184 185 186 187 188 189 190
191 192 193 194 195 196 197 198 244 245;

Hindlimbs - 45 199 200 201 202 203 204 205 206 207 208 209 210 211
212 213 214 215 216 217 218 219 220 221 222 223 224
225 226 227 228 229 230 231 232 233 234 235 236 237
238 239 240 241 242 243;

Ordered characters:

ctype ordered 1 3 8 28 31 43 51 56 67 69-70 72 74 92 107 117 159 168 176 183 193 197
205 213-214 216 219 222 229 233-234;

A.1.4 SCO

Neomorphic-transformational partitions:

neomorphic - 7 4 7 9 16 18 21 24;

transformational - 20 1 2 3 5 6 8 10 11 12 13 14 15 17 19 20 22 23 25 26 27;

3 anatomical partitions:

Extremities - 11 1 2 12 13 15 16 17 18 19 22 24;

Face - 7 3 4 5 6 7 8 20;

Body - 9 9 10 11 14 21 23 25 26 27;

A.1.5 THER

Neomorphic-transformational partitions:

neomorphic - 72 2 12 20 28 29 31 36 47 48 49 51 54 57 58 59 62 71 77
80 81 99 101 116 122 138 149 150 153 156 162 163 164
168 174 176 181 183 190 191 192 202 212 219 220 221 224
226 228 229 236 240 241;

transformational - 190 1 3 4 5 6 7 8 9 10 11 13 14 15 16 17 18 19 21 22 23 24
25 26 27 30 32 33 34 35 37 38 39 40 41 42 43 44 45 46
50 52 53 55 56 60 61 63 64 65 66 67 68 69 70 72 73 74
75 76 78 79 82 83 84 85 86 87 88 89 90 91 92 93 94 95
96 97 98 100 102 103 104 105 106 107 108 109 110 111

112 113 114 115 117 118 119 120 121 123 124 125 126 127
128 129 130 131 132 133 134 135 136 137 139 140 141 142
143 144 145 146 147 148 151 152 154 155 157 158 159 160
161 165 166 167 169 170 171 172 173 175 177 178 179 180
182 184 185 186 187 188 189 193 194 195 196 197 198 199
200 201 203 204 205 206 207 208 209 210 211 213 214 215
216 217 218 222 223 225 227 230 231 232 233 234 235 237
238 239 242;

5 anatomical partitions for THER:

Reproductive (female) - 13 1 2 3 4 5 6 7 8 9 10 11 12 13;

Reproductive (male) - 88 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30
31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47
48 49 50 51 52 53 54 55 56 57 58 59 60 61 62 63 64
65 66 67 68 69 70 71 72 73 74 75 76 77 78 79 80 81
82 83 84 85 86 87 88 89 90 91 92 93 94 95 96 97 98
99 100 101;

Somatic - 99 102 103 104 105 106 107 108 109 110 111 112 113 114
115 116 117 118 119 120 121 122 123 124 125 126 127
128 129 130 131 132 133 134 135 136 137 138 139 140
141 142 143 144 145 146 147 148 149 150 151 152 153
154 155 156 157 158 159 160 161 162 163 164 165 166
167 168 169 170 171 172 173 174 175 176 177 178 179
180 181 182 183 184 185 186 187 188 189 190 191 192
193 194 195 196 197 198 199 200;

Spinnerets - 23 201 202 203 204 205 206 207 208 209 210 211 212 213
214 215 216 217 218 219 220 221 222 223;

Behaviour - 19 224 225 226 227 228 229 230 231 232 233 234 235 236
237 238 239 240 241 242;

Appendix B

Convergence Testing

B.1 Convergence Testing

As discussed in section 3.2.2, several models with a high average standard deviation of split frequencies, indicating incomplete convergence, were rerun under changed settings. Table B.1 contains data on the effects of longer burnin and higher numbers of generations on the average standard deviation of split frequencies (split) and model fit (ML). Model fit usually did not change significantly upon increasing the burnin proportion or the number of generations, though split values are lower. This suggests that the standard models were run for enough generations to obtain reliable marginal likelihoods.

TABLE B.1: Marginal likelihoods (ML) and maximum stepwise average standard deviations of split frequencies (split) of a selection of models. Models were run with the standard burnin proportion and number of generations (25%; 5,000,000), with a higher burnin (HB scheme: 50%; 7,510,000), and with a higher number of generations (MG scheme: 25%; 10,000,000). Both schemes improve convergence but the marginal likelihood remains similar to that of the standard model.

Model	**Standard Model**		**Higher Burnin**		**More Generations**	
	split	ML	split	ML	split	ML
CEAexp - neotrans2	0.0742	−1773.88	0.0695	−1780.71	0.0664	−1774.46
OZLclock - anatomy	0.0900	−3852.92	0.0919	−3849.77	0.0835	−3851.00
THERgamma - IC	0.1463	−3694.09	0.0419	−3697.50	0.0307	−3696.97
CEA - anatomy	0.1003	−1608.51	0.0265	−1608.39	0.0177	−1608.35
HYO - neotrans	0.1233	−2548.43	0.0532	−2548.87	0.0458	−2548.12
SCO - anatomy	0.0940	−259.76	0.0202	−257.92	0.0139	−259.70

Bibliography

Agnarsson, Ingi (2004). "Morphological phylogeny of cobweb spiders and their relatives (Araneae, Araneoidea, Theridiidae)". In: *Zoological Journal of the Linnean Society* 141, pp. 447–626. URL: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.123.8031{\&}rep=rep1{\&}type=pdf.

— (2006). "Asymmetric female genitalia and other remarkable morphology in a new genus of cobweb spiders (Theridiidae, Araneae) from Madagascar". In: *Biological Journal of the Linnean Society* 87.2, pp. 211–232. ISSN: 00244066. DOI: 10.1111/j.1095-8312.2006.00569.x.

Alroy, John (2014). "A simple Bayesian method of inferring extinction". In: *Paleobiology* 40.4, pp. 584–607. ISSN: 0094-8373. DOI: 10.1666/13074.

— (2016). "A simple Bayesian method of inferring extinction: Reply". In: *Ecology* 97.3, pp. 796–798. ISSN: 00129658. DOI: 10.1890/15-0336.1.

Altekar, G. et al. (2004). "Parallel Metropolis-coupled Markov chain Monte Carlo for Bayesian phylogenetic inference". In: *Bioinformatics* 20, pp. 407–415.

Arcila, Dahiana et al. (2015). "An evaluation of fossil tip-dating versus node-age calibrations in tetraodontiform fishes (Teleostei: Percomorphaceae)". In: *Molecular Phylogenetics and Evolution* 82.PA, pp. 131–145. ISSN: 10959513. DOI: 10.1016/j.ympev.2014.10.011. URL: http://dx.doi.org/10.1016/j.ympev.2014.10.011.

Benson, Roger B.J. (2012). "Interrelationships of basal synapsids: Cranial and postcranial morphological partitions suggest different topologies". In: *Journal of Systematic Palaeontology* 10.4, pp. 601–624. ISSN: 14772019. DOI: 10.1080/14772019.2011.631042.

Bielecki, Aleksander et al. (2014). "Diversity of features of the female reproductive system and other morphological characters in leeches (Citellata, Hirudinida) in phylogenetic conception". In: *Cladistics* 30.5, pp. 540–554. ISSN: 10960031. DOI: 10.1111/cla.12058.

Bofkin, Lee and Nick Goldman (2007). "Variation in evolutionary processes at different codon positions". In: *Molecular Biology and Evolution* 24.2, pp. 513–521. ISSN: 07374038. DOI: 10.1093/molbev/msl178.

Brandley, Matthew C, Andreas Schmitz, and Tod W Reeder (2005). "Partitioned Bayesian Analyses, Partition Choice, and the Phylogenetic Relationships of Scincid Lizards". In: *Systematic Biology* 54.3, pp. 373–390. DOI: 10.1080/10635150590946808.

Brazeau, Martin D. (2011). "Problematic character coding methods in morphology and their effects". In: *Biological Journal of the Linnean Society* 104.3, pp. 489–498. ISSN: 00244066. DOI: 10.1111/j.1095-8312.2011.01755.x.

Brazeau, Martin D., Thomas Guillerme, and Martin R. Smith (2019). "An algorithm for Morphological Phylogenetic Analysis with Inapplicable Data". In: *Systematic Biology* 0.0, pp. 0–13.

Brown, Jeremy M. and Alan R. Lemmon (2007). "The importance of data partitioning and the utility of bayes factors in bayesian phylogenetics". In: *Systematic Biology* 56.4, pp. 643–655. ISSN: 10635157. DOI: 10.1080/10635150701546249.

Brown, Jeremy M. and Robert C. Thomson (2018). "The Behavior of Metropolis-Coupled Markov Chains When Sampling Rugged Phylogenetic Distributions". In: *Systematic Biology* 67.4, pp. 729–734. DOI: 10.1093/sysbio/syy008.

Buckley, Thomas R. (2002). "Model misspecification and probabilistic tests of topology: Evidence from empirical data sets". In: *Systematic Biology* 51.3, pp. 509–523. ISSN: 10635157. DOI: 10.1080/10635150290069922.

Bull, J. J. et al. (1993). "Partitioning and Combining Data in Phylogenetic Analysis". In: *Systematic Biology* 42.3, pp. 384–397.

Caccone, Adalgisa et al. (1996). "A molecular phylogeny for the Drosophila melanogaster subgroup and the problem of polymorphism data". In: *Molecular Biology and Evolution* 13.9, pp. 1224–1232. ISSN: 07374038. DOI: 10.1093/oxfordjournals.molbev.a025688.

Cameron, Stephen L. et al. (2012). "A mitochondrial genome phylogeny of termites (Blattodea: Termitoidae): Robust support for interfamilial relationships and molecular synapomorphies define major clades". In: *Molecular Phylogenetics and Evolution* 65.1, pp. 163–173. ISSN: 10557903. DOI: 10.1016/j.ympev.2012.05.034. URL: http://dx.doi.org/10.1016/j.ympev.2012.05.034.

Carlin, Bradley P. and Thomas A. Louis (2000). "Empirical Bayes: Past, Present and Future". In: *Journal of the American Statistical Association* 95.452, pp. 1286–1289.

Casella, George (1985). "An introduction to empirical bayes data analysis". In: *American Statistician* 39.2, pp. 83–87. ISSN: 15372731. DOI: 10.1080/00031305.1985.10479400.

Chang, Jonathan and Michael E. Alfaro (2016). "Crowdsourced geometric morphometrics enable rapid large-scale collection and analysis of phenotypic data". In: *Methods in Ecology and Evolution* 7.4, pp. 472–482. ISSN: 2041210X. DOI: 10.1111/2041-210X.12508.

Chippindale, Paul T. and John J. Wiens (1994). "Weighting, Partitioning, and Combining Characters in Phylogenetic Analysis". In: *Systematic Biology* 43.2, pp. 278–287.

Clarke, Julia A. and Kevin M. Middleton (2008). "Mosaicism, Modules, and the Evolution of Birds: Results from a Bayesian Approach to the Study of Morphological Evolution Using Discrete Character Data". In: *Systematic Biology* 57.2, pp. 185–201. DOI: 10.1080/10635150802022231.

Clarke, Julia A. and Mark A. Norell (2002). "The Morphology and Phylogenetic Position of Apsaravis ukhaana from the Late Cretaceous of Mongolia". In: *American Museum Novitates* 3387.3387, pp. 1–46. ISSN: 0003-0082. DOI: 10.1206/0003-0082(2002)387<0001:tmappo>2.0.co;2.

Clarke, Julia A., Zhonghe Zhou, and Fucheng Zhang (2006). "Insight into the evolution of avian flight from a new clade of Early Cretaceous ornithurines from China and the morphology of Yixianornis grabaui". In: *Journal of Anatomy* 208, pp. 287–308.

Crick, F. H. C. (1966). "Codon-Anticodon pairing: The Wobble Hypothesis". In: *Journal of Molecular Biology* 19, pp. 548–555.

Dahm, Ralf (2008). "Discovering DNA: Friedrich Miescher and the early years of nucleic acid research". In: *Human Genetics* 122.6, pp. 565–581. ISSN: 03406717. DOI: 10.1007/s00439-007-0433-0.

Dayhoff, Margaret O. (1965). "Computer aids to protein sequence determination". In: *Journal of Theoretical Biology* 8.1, pp. 97–112.

— (1969). *Atlas of Protein Sequence and Structure, Volume 4*.

De Queiroz, Alan (1993). "For consensus (sometimes)". In: *Systematic Biology* 42.3, pp. 368–372. ISSN: 1076836X. DOI: 10.1093/sysbio/42.3.368.

Derome, Nicolas et al. (2002). "Phylogeny of Antarctic dragonfishes (Bathydraconidae, Notothenioidei, Teleostei) and related families based on their anatomy and two mitochondrial genes". In: *Molecular Phylogenetics and Evolution* 24.1, pp. 139–152. ISSN: 10557903. DOI: 10.1016/S1055-7903(02)00223-3.

Deverreux, Daniel and Pierre Pellegrin (1990). *Biologie, Logique et Métaphysique chez Aristote.* Ed. by D Deverreux and P Pellegrin. Paris: Éditions du Centre National de la Recherche Scientifique.

Diogo, Rui, Julia L. Molnar, and Bernard Wood (2017). "Bonobo anatomy reveals stasis and mosaicism in chimpanzee evolution, and supports bonobos as the most appropriate extant model for the common ancestor of chimpanzees and humans". In: *Scientific Reports* 7.1, pp. 1–8. ISSN: 20452322. DOI: 10.1038/s41598-017-00548-3. URL: http://dx.doi.org/10.1038/s41598-017-00548-3.

Dos Reis, Mario, Philip C. J. Donoghue, and Ziheng Yang (2016). "Bayesian molecular clock dating of species divergences in the genomics era". In: *Nature Reviews Genetics* 17.2, pp. 71–80. ISSN: 14710064. DOI: 10.1038/nrg.2015.8.

Duchêne, Sebastián et al. (2011). "Mitogenome phylogenetics: The impact of using single regions and partitioning schemes on topology, substitution rate and divergence time estimation". In: *PLoS ONE* 6.11. ISSN: 19326203. DOI: 10.1371/journal.pone.0027138.

Edwards, W. F. and Luigi Luca Cavalli-Sforza (1963). "The reconstruction of evolution". In: *Annual Review of Human Genetics* 27.105.

Ekman, Stefan and Rakel Blaalid (2011). "The devil in the details: Interactions between the branch-length prior and likelihood model affect node support and

branch lengths in the phylogeny of the psoraceae". In: *Systematic Biology* 60.4, pp. 541–561. ISSN: 10635157. DOI: 10.1093/sysbio/syr022.

Engel, Michael S. and David A. Grimaldi (2007). "Cretaceous Scolebythidae and Phylogeny of the Family (Hymenoptera: Chrysidoidea)". In: *American Museum Novitates* 3568.1, p. 1. ISSN: 0003-0082. DOI: 10.1206/0003-0082(2007)475[1:csapot]2.0.co;2.

Engel, Michael S., Jaime Ortega-Blanco, and Ryan C. McKellar (2013). "New scolebythid wasps in Cretaceous amber from Spain and Canada, with implications for the phylogeny of the family (Hymenoptera: Scolebythidae)". In: *Cretaceous Research* 46, pp. 31–42. ISSN: 01956671. DOI: 10.1016/j.cretres.2013.09.003. URL: http://dx.doi.org/10.1016/j.cretres.2013.09.003.

Faith, Daniel P. and John W. H. Trueman (2001). "Towards an Inclusive Philosophy for Phylogenetic Inference". In: *Systematic Biology* 50.3, pp. 331–350. ISSN: 10635157. DOI: 10.1080/106351501300317969.

Farris, James S. (1969). "A Successive Approximations Approach to Character Weighting". In: *Systematic Zoology* 18.4, p. 374. ISSN: 00397989. DOI: 10.2307/2412182.

Felice, Ryan N. and Anjali Goswami (2018). "Developmental origins of mosaic evolution in the avian cranium". In: *Proceedings of the National Academy of Sciences of the United States of America* 115.3, pp. 555–560. ISSN: 10916490. DOI: 10.1073/pnas.1716437115.

Felsenstein, Joseph (1978a). "Cases in which Parsimony or Compatibility Methods will be Positively Misleading". In: *Systematic Biology* 27.4, pp. 401–410. ISSN: 1076836X. DOI: 10.1093/sysbio/27.4.401.

— (1978b). "The number of evolutionary trees". In: *Systematic Zoology* 27.1, pp. 71–78. ISSN: 00397989. DOI: 10.2307/2412810.

— (2001). *Inferring Phylogenies*.

Field, Daniel J. et al. (2018). "Early Evolution of Modern Birds Structured by Global Forest Collapse at the End-Cretaceous Mass Extinction". In: *Current Biology* 28.11, 1825–1831.e2. ISSN: 09609822. DOI: 10.1016/j.cub.2018.04.062. URL: https://doi.org/10.1016/j.cub.2018.04.062.

Fourment, Mathieu et al. (2019). "19 Dubious Ways To Compute the Marginal Likelihood of a Phylogenetic Tree Topology". In: *Systematic Biology* 0.0, pp. 1–12. DOI: 10.1093/sysbio/syz046. arXiv: 1811.11804. URL: http://arxiv.org/abs/1811.11804.

Gamble, Tony et al. (2012). "Repeated origin and loss of adhesive toepads in Geckos". In: *PLoS ONE* 7.6. ISSN: 19326203. DOI: 10.1371/journal.pone.0039429.

Gaubert, Philippe et al. (2005). "Mosaics of convergences and noise in morphological phylogenies: what's in a viverrid-like carnivoran?" In: *Systematic biology* 54.6, pp. 865–894. ISSN: 10635157. DOI: 10.1080/10635150500232769.

Gauthier, Jacques A., Arnold G. Kluge, and Timothy Rowe (1988). "Amniote Phylogeny and the Importance of Fossils". In: *Cladistics* 4.2, pp. 105–209. ISSN: 10960031. DOI: 10.1111/j.1096-0031.1988.tb00514.x.

Geyer, Charles J. (1992). "Practical Markov Chain Monte Carlo". In: *Statistical Science* 7.4, pp. 473–483. DOI: 10.1097/EDE.0b013e3181. URL: https://www.jstor.org/stable/2246094.

Göker, Markus and Guido W. Grimm (2008). "General functions to transform associate data to host data, and their use in phylogenetic inference from sequences with intra-individual variability". In: *BMC Evolutionary Biology* 8.1, pp. 1–24. ISSN: 14712148. DOI: 10.1186/1471-2148-8-86.

Goloboff, Pablo A. (1993). "Estimating Character weights during tree search". In: *Cladistics* 9, pp. 83–91.

— (2013). "Extended implied weighting". In: *Cladistics* 1, pp. 1–13.

Goloboff, Pablo A. et al. (2008). "Weighting against homoplasy improves phylogenetic analysis of morphological data sets". In: *Cladistics* 24.2008, pp. 1–16.

Goloboff, Pablo A. et al. (2018). "Morphological Data Sets Fit a Common Mechanism Much More Poorly than DNA Sequences and Call Into Question the Mkv Model". In: *Systematic Biology* 0.0, pp. 1–11.

Gontier, Nathalie (2011). "Depicting the Tree of Life: The Philosophical and Historical Roots of Evolutionary Tree Diagrams". In: *Evolution: Education and Outreach* 4.3, pp. 515–538. ISSN: 19366434. DOI: 10.1007/s12052-011-0355-0.

Graybeal, Anna (1998). "Is It better to add taxa or characters to a difficult phylogenetic problem ?" In: *Systematic Biology* 47.1, pp. 9–17.

Green, Edwin J., William E. Strawderman, and Charles E. Thomas (1992). "Empirical Bayes Development of Honduran Pine Yield Models". In: *Forest Science* 38.1, pp. 21–33.

Hayes, Reilly F. et al. (2020). "Modeling the dynamics of a Late Triassic vertebrate extinction: The Adamanian/Revueltian faunal turnover, Petrified Forest National Park, Arizona, USA". In: *Geology* 48.4, pp. 318–322. ISSN: 0091-7613. DOI: 10.1130/g47037.1.

Hillis, David M., John P. Huelsenbeck, and David L. Swofford (1994). "Hobgoblin of phylogenetics?" In: *Nature* 369.6479, pp. 363–364. ISSN: 00280836. DOI: 10.1038/369363a0.

Holder, Mark and Paul O. Lewis (2003). "Phylogeny estimation: traditional and Bayesian approaches". In: *Nature Reviews Genetics* 4, pp. 175–284.

Hoyal Cuthill, Jennifer F., Simon J. Braddy, and Philip C. J. Donoghue (2010). "A formula for maximum possible steps in multistate characters: Isolating matrix parameter effects on measures of evolutionary convergence". In: *Cladistics* 26.1, pp. 98–102. ISSN: 07483007. DOI: 10.1111/j.1096-0031.2009.00270.x.

Huelsenbeck, John P. (1991). "When are fossils better than extant taxa in phylogenetic analysis?" In: *Systematic Biology* 40.4, pp. 458–469. ISSN: 1076836X. DOI: 10.1093/sysbio/40.4.458.

— (1994). "Comparing the stratigraphic record to estimates of phylogeny". In: *Paleobiology* 20.4, pp. 470–483.

Huelsenbeck, John P. (1995). "Performance of phylogenetic methods in simulation". In: *Systematic Biology* 44.1, pp. 17–48. ISSN: 1076836X. DOI: 10.1093/sysbio/44.1.17.

— (1997). "Is the Felsenstein zone a fly trap?" In: *Systematic Biology* 46.1, pp. 69–74. ISSN: 10635157. DOI: 10.1093/sysbio/46.1.69.

Huelsenbeck, John P. and Fredrik Ronquist (2001). "MRBAYES: Bayesian inference of phylogenetic trees". In: *Bioinformatics* 17.8, pp. 754–755. ISSN: 13674803. DOI: 10.1093/bioinformatics/17.8.754.

Janzen, Daniel H. (2009). "A DNA barcode for land plants". In: *Proceedings of the National Academy of Sciences of the United States of America* 106.31, pp. 12794–12797. ISSN: 17550998. DOI: 10.1111/1755-0998.12194.

Johnson, Walter E. and Robert K. Selander (1971). "Protein variation and systematics in kangaroo rats (Genus Dipodomys)". In: *Systematic Zoology* 20, pp. 377–405.

Jukes, T.H. and C.R. Cantor (1969). "Evolution of protein molecules". In: *Mammalian Protein Metabolism*. Ed. by H.N. Munro. New York: Academic Press, pp. 21–132.

Kainer, David and Robert Lanfear (2015). "The effects of partitioning on phylogenetic inference". In: *Molecular Biology and Evolution* 32.6, pp. 1611–1627. ISSN: 15371719. DOI: 10.1093/molbev/msv026.

Källersjö, M., Victor A. Albert, and James S. Farris (1999). "Homoplasy increases phylogenetic structure". In: *Cladistics* 15.1, pp. 91–93. ISSN: 07483007. DOI: 10.1006/clad.1999.0085.

Kass, Robert E. and Adrian E. Raftery (1995). "Bayes factors". In: *Journal of the American Statistical Association* 90.430, pp. 773–795.

Kawahara, Akito Y. et al. (2017). "A molecular phylogeny and revised higher-level classification for the leaf-mining moth family Gracillariidae and its implications for larval host-use evolution". In: *Systematic Entomology* 42.1, pp. 60–81. ISSN: 13653113. DOI: 10.1111/syen.12210.

Kearney, Maureen and James M. Clark (2003). "Problems due to missing data in phylogenetic analyses including fossils: A critical review". In: *Journal of Vertebrate Palaeontology* 23.2, pp. 263–274.

Kimura, Motoo (1983). *The neutral theory of molecular evolution*. Cambridge University Press, p. 367. ISBN: 9780511623486. DOI: 10.1017/cbo9780511623486.004.

Klug, A. (1968). "Rosalind Franklin and the discovery of the structure of DNA". In: *Nature* 219.5156, pp. 808–844. ISSN: 00280836. DOI: 10.1038/219808a0.

Kluge, Arnold G. (1989). "A concern for evidence and a phylogenetic hypothesis of relationships among Epicrates (Boidae, Serpentes)". In: *Systematic Zoology* 38.1, pp. 7–25.

— (1997). "Sophisticated falsification and research cycles: Consequences for differential character weighting in phylogenetic systematics". In: *Zoologica Scripta* 26.4, pp. 349–360. ISSN: 03003256. DOI: 10.1111/j.1463-6409.1997.tb00424.x.

Krause, David W. et al. (2014). "First cranial remains of a gondwanatherian mammal reveal remarkable mosaicism". In: *Nature* 515.7528, pp. 512–517. ISSN: 14764687. DOI: 10.1038/nature13922. URL: http://dx.doi.org/10.1038/nature13922.

Lanciotti, R. S. et al. (1999). "Origin of the West Nile virus responsible for an outbreak of encephalitis in the Northeastern United States". In: *Science* 286.5448, pp. 2333–2337. ISSN: 00368075. DOI: 10.1126/science.286.5448.2333.

Lanfear, Robert et al. (2012). "PartitionFinder: Combined selection of partitioning schemes and substitution models for phylogenetic analyses". In: *Molecular Biology and Evolution* 29.6, pp. 1695–1701. ISSN: 07374038. DOI: 10.1093/molbev/mss020.

Lanfear, Robert et al. (2016). "PartitionFinder 2: New Methods for Selecting Partitioned Models of Evolution for Molecular and Morphological Phylogenetic Analyses". In: *Molecular Biology and Evolution* 34.3, pp. 772–773. DOI: 10.1093/molbev/msw260.

Lee, Michael S. Y. and Alessandro Palci (2015). "Morphological phylogenetics in the genomic age". In: *Current Biology* 25.19, R922–R929. ISSN: 09609822. DOI: 10.1016/j.cub.2015.07.009. URL: http://dx.doi.org/10.1016/j.cub.2015.07.009.

Lee, Michael S. Y. and Trevor H. Worthy (2012). "Likelihood reinstates Archaeopteryx as a primitive bird". In: *Biology Letters* 8.2, pp. 299–303. ISSN: 1744957X. DOI: 10.1098/rsbl.2011.0884.

Lee, Michael S. Y. et al. (2014). "Morphological clocks in paleontology, and a mid-cretaceous origin of crown aves". In: *Systematic Biology* 63.3, pp. 442–449. ISSN: 1076836X. DOI: 10.1093/sysbio/syt110.

Lewis, Paul O. (2001). "A Likelihood Approach to Estimating Phylogeny from Discrete Morphological Character Data". In: *Systematic Biology* 50.6, pp. 913–925. DOI: 10.1080/106351501753462876.

Löytynoja, Ari and Nick Goldman (2008). "Phylogeny-aware gap placement prevents errors in sequence alignment and evolutionary analysis". In: *Science* 320.5883, pp. 1632–1635. ISSN: 00368075. DOI: 10.1126/science.1158395.

Maddison, David R., David L. Swofford, and Wayne P. Maddison (1997). "NEXUS: An Extensible File Format for Systematic Information". In: *Systematic Biology* 46.4, pp. 590–621. ISSN: 10635157. DOI: 10.2307/2413497.

Manceau, Marie et al. (2010). "Convergence in pigmentation at multiple levels: Mutations, genes and function". In: *Philosophical Transactions of the Royal Society B: Biological Sciences* 365.1552, pp. 2439–2450. ISSN: 14712970. DOI: 10.1098/rstb.2010.0104.

Marshall, David C., Chris Simon, and Thomas R. Buckley (2006). "Accurate branch length estimation in partitioned Bayesian analyses requires accommodation of among-partition rate variation and attention to branch length priors". In: *Systematic Biology* 55.6, pp. 993–1003. ISSN: 10635157. DOI: 10.1080/10635150601087641.

McGuire, Jimmy A. et al. (2007). "Phylogenetic Systematics and Biogeography of Hummingbirds: Bayesian and Maximum Likelihood Analyses of Partitioned Data and Selection of an Appropriate Partitioning Strategy". In: *Systematic Biology* 56.5, pp. 837–856.

Mickevich, M. F. (1978). "Taxonomic Congruence". In: *Systematic Zoology* 27, pp. 143–158. ISSN: 00397989. DOI: 10.2307/2412969.

Miller, Robert L. and Everett C. Olson (1960). "Morphological Integration: A discussion". In: 14.1, pp. 132–133.

Moorthie, Sowmiya, Christopher J. Mattocks, and Caroline F. Wright (2011). "Review of massively parallel DNA sequencing technologies". In: *HUGO Journal* 5.1-4, pp. 1–12. ISSN: 18776558. DOI: 10.1007/s11568-011-9156-3.

Morgan, Gregory J. (1998). "Emile Zuckerkandl, Linus Pauling, and the Molecular Evolutionary Clock, 1959-1965". In: *Journal of the History of Biology* 31.2, pp. 155–178.

Moysiuk, Joseph and Jean-Bernard Caron (2019). "A new hurdiid radiodont from the Burgess Shale evinces the exploitation of Cambrian infaunal food sources". In: *Proceedings of the Royal Society B: Biological Sciences* 286.1908, p. 20191079. ISSN: 0962-8452. DOI: 10.1098/rspb.2019.1079. URL: https://royalsocietypublishing.org/doi/10.1098/rspb.2019.1079.

Nylander, Johan A. A. et al. (2004). "Bayesian Phylogenetic Analysis of Combined Data". In: *Systematic Biology* 53.1, pp. 47–67. ISSN: 10635157. DOI: 10.1080/10635150490264699.

Oaks, Jamie R. et al. (2019). "Marginal likelihoods in phylogenetics: a review of methods and applications". In: pp. 1–51.

O'Connor, Jingmai K. and Zhonghe Zhou (2013). "A redescription of chaoyangia beishanensis (aves) and a comprehensive phylogeny of mesozoic birds". In: *Journal of Systematic Palaeontology* 11.7, pp. 889–906. ISSN: 14772019. DOI: 10.1080/14772019.2012.690455.

O'Leary, Maureen A. et al. (2013). "The Placental Mammal Ancestor and the Post-K-Pg Radiation of Placentals". In: *Science* 339.6120, pp. 662–667. DOI: 10.1126/science.1229237.

Pagel, Mark and Francois Lutzoni (2002). *Accounting for phylogenetic uncertainty in comparative studies of evolution and adaptation*.

Paradis, Emmanuel et al. (2019). *Package 'ape': Analyses of Phylogenetics and Evolution*. URL: https://orcid.org/0000-0002-2127-0443.

Parins-Fukuchi, Caroline (2018). "Use of Continuous Traits Can Improve Morphological Phylogenetics". In: *Systematic Biology* 67.2, pp. 328–339. ISSN: 1076836X. DOI: 10.1093/sysbio/syx072.

Polasky, Stephen et al. (2001). "A comparison of taxonomic distinctness versus richness as criteria for setting conservation priorities for North American birds". In: *Biological Conservation* 97.1, pp. 99–105. ISSN: 00063207. DOI: 10.1016/S0006-3207(00)00103-8.

Posada, David and Keith A. Crandall (2001). "Selecting the Best-Fit Model of Nucleotide Substitution". In: *Systematic Biology* 50.4, pp. 580–601. ISSN: 10635157. DOI: 10.1080/106351501750435121.

Potter, D. et al. (2007). *Phylogeny and classification of Rosaceae*. Vol. 266. 1-2, pp. 5–43. ISBN: 0060600705. DOI: 10.1007/s00606-007-0539-9.

Poux, Céline et al. (2008). "Molecular phylogeny and divergence times of Malagasy tenrecs: Influence of data partitioning and taxon sampling on dating analyses". In: *BMC Evolutionary Biology* 8.1, pp. 1–16. ISSN: 14712148. DOI: 10.1186/1471-2148-8-102.

Ragan, Mark A. (2009). "Trees and networks before and after Darwin". In: *Biology Direct* 4, pp. 319–329. ISSN: 17456150. DOI: 10.1186/1745-6150-4-43.

Rannala, Bruce, Tianqi Zhu, and Ziheng Yang (2012). "Tail paradox, partial identifiability, and influential priors in bayesian branch length inference". In: *Molecular Biology and Evolution* 29.1, pp. 325–335. ISSN: 07374038. DOI: 10.1093/molbev/msr210. URL: https://academic.oup.com/mbe/article-lookup/doi/10.1093/molbev/msr210.

Robinson, D. F. and L. R. Foulds (1981). "Comparison of phylogenetic trees". In: *Mathematical Biosciences* 53.1-2, pp. 131–147. ISSN: 00255564. DOI: 10.1016/0025-5564(81)90043-2.

Rodrigo, Allen G. et al. (1993). "A randomisation test of the null hypothesis that two cladograms are sample estimates of a parametric phylogenetic tree". In: *New Zealand Journal of Botany* 31.3, pp. 257–268. ISSN: 11758643. DOI: 10.1080/0028825X.1993.10419503.

Rodrigues, Alexandre et al. (2019). "Reducing power companies billing costs via empirical bayes and seasonality remover". In: *Engineering Applications of Artificial Intelligence* 81.September 2018, pp. 387–396. ISSN: 09521976. DOI: 10.1016/j.engappai.2019.01.007. URL: https://doi.org/10.1016/j.engappai.2019.01.007.

Ronquist, Fredrik, John P. Huelsenbeck, and Maxim Teslenko (2011). *Mrbayes3.2 Manual*. URL: http://mrbayes.sourceforge.net/mb3.2{_}manual.pdf.

Ronquist, Fredrik, Nicolas Lartillot, and Matthew J. Phillips (2016). "Closing the gap between rocks and clocks using total-evidence dating". In: *Philosophical Transactions of the Royal Society B: Biological Sciences* 371.1699. ISSN: 14712970. DOI: 10.1098/rstb.2015.0136.

Ronquist, Fredrik et al. (2012). "Mrbayes 3.2: Efficient bayesian phylogenetic inference and model choice across a large model space". In: *Systematic Biology* 61.3, pp. 539–542. ISSN: 10635157. DOI: 10.1093/sysbio/sys029.

Rønsted, Nina et al. (2012). "Can phylogeny predict chemical diversity and potential medicinal activity of plants? A case study of amaryllidaceae". In: *BMC Evolutionary Biology* 12.1, pp. 24–29. ISSN: 14712148. DOI: 10.1186/1471-2148-12-182.

Rosa, Brunno B, Gabriel A R Melo, and Marcos S Barbeitos (2019). "Homoplasy-based partitioning outperforms alternatives in Bayesian analysis of discrete morphological data". In: *Systematic Biology* 68.4, pp. 657–671. ISSN: 1063-5157. DOI: 10.1093/sysbio/syz001. URL: http://dx.doi.org/10.1093/sysbio/syz001.

Rubino, Gerardo and Bruno Tuffin (2009). *Rare Event Simulation using Monte Carlo Methods*. John Wiley & Sons, p. 278.

Sanger, F. and A. R. Coulson (1975). "A rapid method for determining sequences in DNA by primed synthesis with DNA polymerase". In: *Journal of Molecular Biology* 94.3, pp. 441–448. ISSN: 00222836. DOI: 10.1016/0022-2836(75)90213-2.

Sanger, F., S. Nicklen, and A. R. Coulson (1977). "DNA sequencing with chain-terminating inhibitors". In: *Proceedings of the National Academy of Sciences* 74.12, pp. 5463–5467. ISSN: 1873233X. DOI: 10.1097/00006250-199004001-00013.

Saslis-Lagoudakis, C. Haris et al. (2012). "Phylogenies reveal predictive power of traditional medicine in bioprospecting". In: *Proceedings of the National Academy of Sciences of the United States of America* 109.39, pp. 15835–15840. ISSN: 00278424. DOI: 10.1073/pnas.1202242109.

Schaap, Pauline et al. (2006). "Molecular phylogeny and evolution of morphology in the social amoebas". In: *Science* 314.5799, pp. 661–663. ISSN: 00368075. DOI: 10.1126/science.1130670.

Schleicher, August (1853). "Die ersten Spaltungen des indogermanischen Urvolkes." In: *Allgemeine Monatsschrift fuer Wissenschaft und Literatur* 3, pp. 786–787.

Schnell, Gary D., Troy L. Best, and Michael L. Kennedy (1978). "Interspecific morphologic variation in kangaroo rats (dipodomys): Degree of concordance with genic variation". In: *Systematic Zoology* 27.1, pp. 34–48. ISSN: 00397989. DOI: 10.2307/2412811.

Sereno, Paul C. (2007). "Logical basis for morphological characters in phylogenetics". In: *Cladistics* 23.6, pp. 565–587. ISSN: 07483007. DOI: 10.1111/j.1096-0031.2007.00161.x.

Smith, Martin R. (2019a). "Bayesian and parsimony approaches reconstruct informative trees from simulated morphological datasets". In: *Biology Letters* 15.2. ISSN: 1744957X. DOI: 10.1098/rsbl.2018.0632.

— (2019b). *Interpreting large quartet distances*.

— (2019c). *Phylogenetic tree search using custom optimality criteria*. DOI: 10.1093/sysbio/syy083.

— (2019d). *Quartet: comparison of phylogenetic trees using quartet and bipartition measures*. DOI: 10.5281/zenodo.2536318.

— (2019e). *Tree Distance Metrics*.

Sokal, R. R. and Charles D. Michener (1967). "The effects of different numerical techniques on the phenetic classifications of bees of the Hoplitis complex (Megachilidae). Proc. Linn. Soc. Lond. 178:54-74." In: *Proceedings of the Linnean Society* 178, pp. 54–74.

Solow, Andrew R. (2016). "A simple Bayesian method of inferring extinction: Comment". In: *Ecology* 97.3, pp. 796–798. ISSN: 00129658. DOI: 10.1890/15-0336.1.

Springer, Mark S. et al. (2013). "Technical comment on "The placental mammal ancestor and the post-K-Pg radiation of placentals"". In: *Science* 341.6146. ISSN: 10959203. DOI: 10.1126/science.1238025.

Steel, Mike A. and David Penny (2006). "Maximum parsimony and the phylogenetic information in multistate characters". In: *Parsimony, Phylogeny, and Genomics*. Ed. by Victor A. Albert, pp. 163–178.

Strugnell, Jan et al. (2005). "Molecular phylogeny of coleoid cephalopods (Mollusca: Cephalopoda) using a multigene approach; the effect of data partitioning on resolving phylogenies in a Bayesian framework". In: *Molecular Phylogenetics and Evolution* 37.2, pp. 426–441. ISSN: 10557903. DOI: 10.1016/j.ympev.2005.03.020.

Sun, Haijing et al. (2018a). *Electronic supplementary material for: Hyoliths with pedicles illuminate the origin of the brachiopod body plan*. URL: https://ms609.github.io/hyoliths/bayesian.html.

— (2018b). "Hyoliths with pedicles illuminate the origin of the brachiopod body plan". In: *Proceedings of the Royal Society B: Biological Sciences* 285.20181780. ISSN: 0962-8452. DOI: 10.1098/rspb.2018.1780. URL: https://royalsocietypublishing.org/doi/10.1098/rspb.2018.1780.

Swenson, Ulf, James E. Richardson, and Igor V. Bartish (2008). "Multi-gene phylogeny of the pantropical subfamily Chrysophylloideae (Sapotaceae): evidence of generic polyphyly and extensive morphological homoplasy". In: *Cladistics* 24, pp. 1006–1031.

Tarasov, Sergei and Francois Génier (2015). "Innovative bayesian and parsimony phylogeny of dung beetles (coleoptera, scarabaeidae, scarabaeinae) enhanced by ontology-based partitioning of morphological characters". In: *PLoS ONE* 10.3. ISSN: 19326203. DOI: 10.1371/journal.pone.0116671.

Thompson, Richard S. et al. (2012). "Phylogeny of the ankylosaurian dinosaurs (Ornithischia: Thyreophora)". In: *Journal of Systematic Palaeontology* 10.2, pp. 301–312. ISSN: 14772019. DOI: 10.1080/14772019.2011.569091.

Verboon, Annemieke R. (2014). "The Medieval Tree of Porphyry: An Organic Structure of Logic". In: *The Tree: SYmbol, Allegory, and Mnemonic Device in Medieval Art and Thought*. Turnhout: Brepols: Brepols Publishers, pp. 95–116.

Vogt, Lars (2018). "The logical basis for coding ontologically dependent characters". In: *Cladistics* 34.4, pp. 438–458. ISSN: 10960031. DOI: 10.1111/cla.12209.

Wiens, John J. (2000). *Phylogenetic analysis of morphological data*. Ed. by John J. Wiens. Washington, D.C.: Smithsonian Institution Press.

Wilkins, John S. (2009). *Defining species: A sourcebook from antiquity to today*, p. 224.

Wilkinson, Mark (1995). "Coping with abundant missing entries in phylogenetic inference using parsimony". In: *Systematic Biology* 44.4, pp. 501–514. ISSN: 1076836X. DOI: 10.1093/sysbio/44.4.501.

Wright, April Marie (2015). "Estimating phylogenetic trees from discrete morphological data". PhD thesis, pp. 1–114.

Wu, Shu, Jie Xiong, and Yuhe Yu Yu (2015). "Taxonomic resolutions based on 18S rRNA Genes: A case study of subclass Copepoda". In: *PLoS ONE* 10.6, pp. 1–19. ISSN: 19326203. DOI: 10.1371/journal.pone.0131498.

Xie, Wangang et al. (2011). "Improving Marginal Likelihood Estimation for Bayesian Phylogenetic Model Selection". In: *Systematic Biology* 60.2, pp. 150–160. DOI: 10.1093/sysbio/syq085.

Yang, Ziheng (1996). "Among-site rate variation and its impact on phylogenetic analyses". In: *Trends in Ecology and Evolution* 11.9, pp. 367–372. ISSN: 01695347. DOI: 10.1016/0169-5347(96)10041-0.

Yates, Adam M. (2007). "Solving a dinosaurian puzzle: the identity of Aliwalia rex Galton". In: *Historical Biology* 19.1, pp. 93–123.

Zhang, Chi, Bruce Rannala, and Ziheng Yang (2012). "Robustness of compound dirichlet priors for bayesian inference of branch lengths". In: *Systematic Biology* 61.5, pp. 779–784. ISSN: 10635157. DOI: 10.1093/sysbio/sys030.